HPE DSI Town Hall Meeting

October 23, 2020

UNIVERSITY of
HOUSTON
DIVISION OF RESEARCH

Mission, Vision

- **Vision:** Delivering research, education and services in data science and scientific computing to the Houston metro area and beyond

- **Mission:** The University of Houston's Hewlett Packard Enterprise Data Science Institute, in collaboration with departments and colleges, leads research, education, and service activities in the broad area of data science and scientific computing and their applications.

Goals

- **Collaborative programs** with public and private sector partners to advance data science and scientific computing in the metro area and beyond

- **Research collaborations** to increase the competitiveness of UH faculty for research grants

- **Education of the current and future data science workforce** through degree programs, non-degree certificate programs, hands-on experiences, and workshops

- **Community-based programs** for middle and high school students in partnership with colleges

- **Resources and services** in the form of expertise, hardware, and software

# External Advisory Board

- External Advisory Board
  - Maury Blackman ('92), Chief Executive Officer, Premise
  - Rashmi Kumar, Senior Vice President and Chief Information Officer at Hewlett Packard Enterprise
  - Gregory Leveille, Chief Technology Officer, Conoco Phillips
  - Tray E. Cockerell, Strategic Relationships Executive, Humana Inc.
  - Mark Kuvshinikov (JD, '09), Partner, McKinsey & Company
  - Rustom Mody (MBA, '86), Former Vice President of Technical Excellence, Baker Hughes, a GE Company

Faculty Interest Groups

- 30 faculty across 7 colleges

- Open to anyone interested

- Areas
  - Foundations
  - Applications
  - Industry Collaborations
  - Education
  - Core Facility

2. Application Areas	a. Social Sciences
	b. Arts and Humanities
	c. Engineering
	d. Physical Sciences
	e. Life Sciences
	f. Health
	g. Energy/Power
	h. Materials
	i. Law
	j. Business
	k. Technology
	l. Policy
	m. Cyber Physical Systems/Cyber Security

Organizational Chart
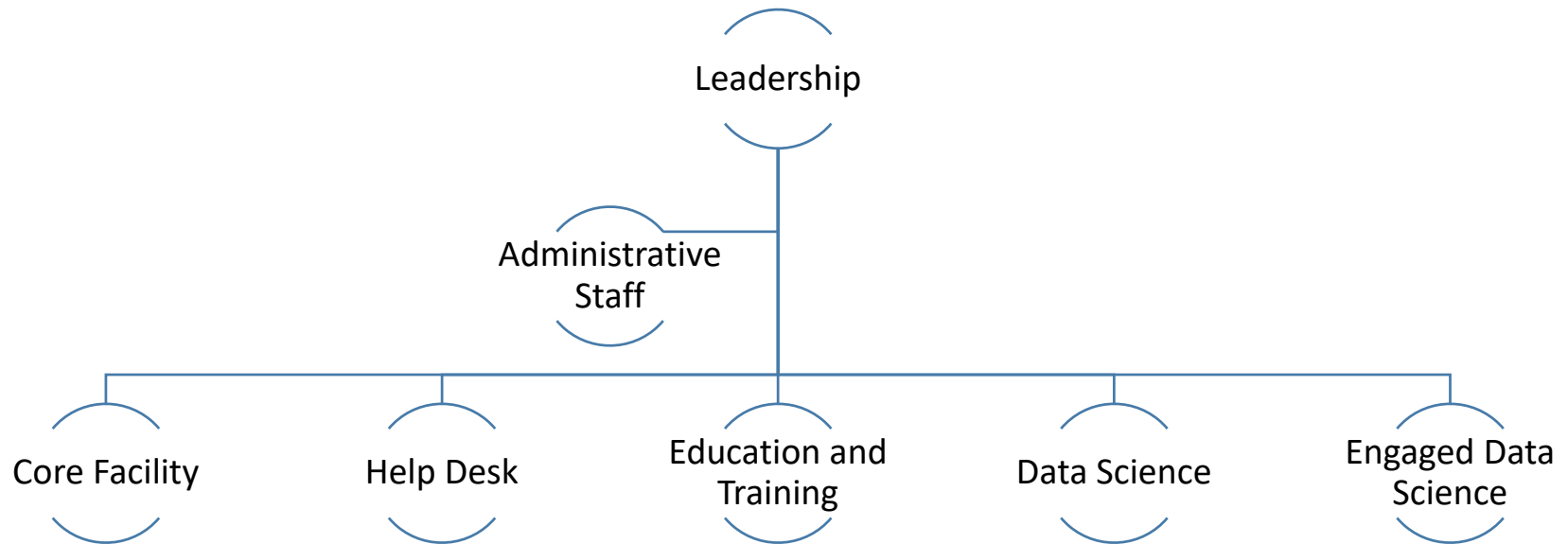
Staff

- Claudia Neuhauser, PhD, Director
- Martin Huarte-Espinosa, PhD, Associate Director
- Andrew Kapral, Director of Engaged Data Science
- Jerry Ebalunode, PhD, Senior Researcher
- Ishita Sharma, Data Scientist
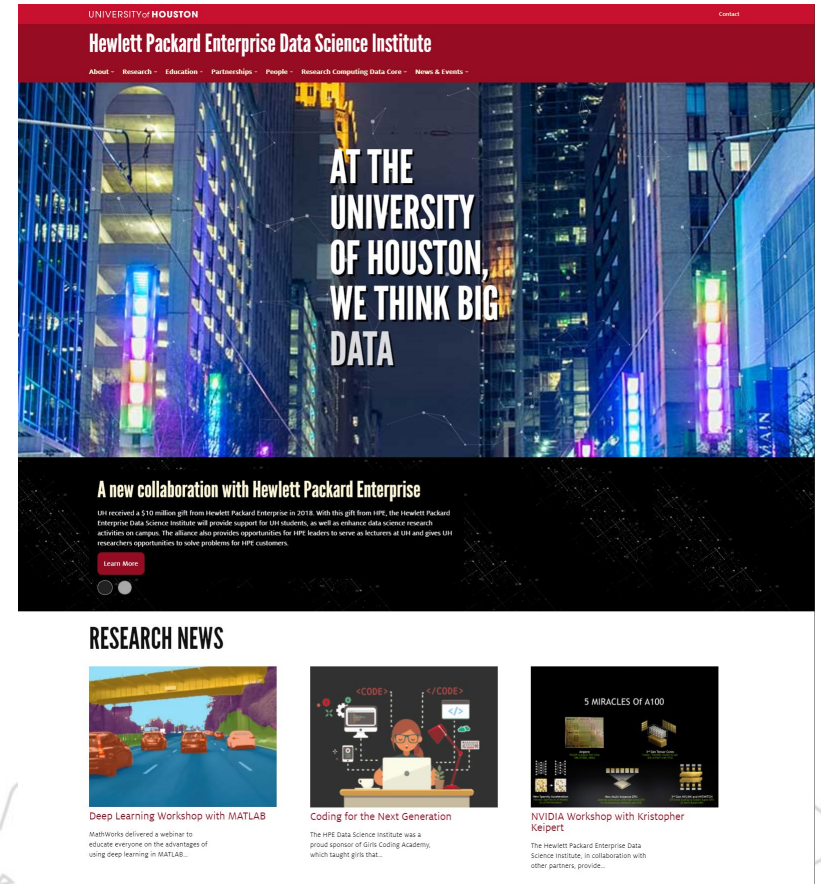- Helaine Guillory, Administration

New Staff

- Ishita Sharma
  - Data Scientist
  - M.S. Computer and Information Science, Texas A&M
  - B.E. Electronics and Communication, Punjab University, India

- Projects
  - Humana
  - Education

New Website

- ## https://hpedsi.uh.edu/
  - Research News
  - Events
  - Policies
  - Courses
  - Resources
  - Staff
  - …

Events

FEATURED EVENTS

October 22 — 11:30am - 1:00pm
HPE DSI Seminar Series 2020
Accelerating Medicine…
With Mark Oldakowski, Chief Operating Officer, Bionano Genomics, Inc., in partnership with the Society for HPC Professionals

October 23 — 2:00pm - 3:00pm
Hewlett Packard Enterprise Data Science Institute University of Houston
HPE Data Science Institute
Town Hall Meeting
A meeting with Claudia Neuhauser, Director, and Martin Huarte Espinosa, Associate Director, to update you about recent activities in the four broad areas the HPE Data Science Institute is engaging in and to seek your feedback.

October 28 — 11:00am - 12:00pm
HPE DSI Seminar Series 2020
Weaker the Better
With Bas Peters of Emory University, in partnership with the Geophysical Society of Houston

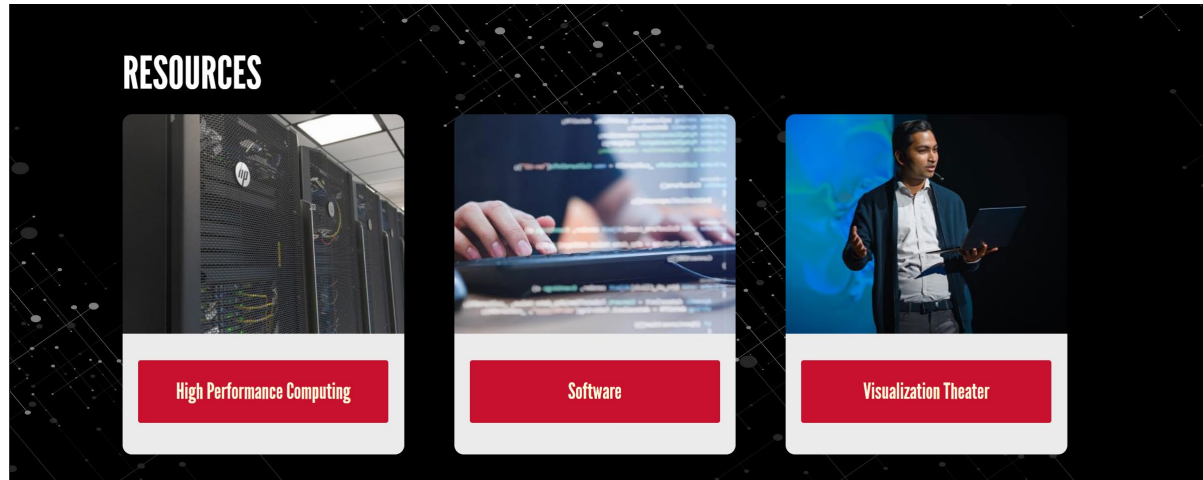- Register on our website for virtual events
  - Partnerships
    - Society for HPC Professionals
    - Geophysical Society of Houston

UNIVERSITY of
HOUSTON
DIVISION OF RESEARCH

Resources

Martin Huarte-Espinosa

Resources

- High Performance Computing

- Software

- Visualization Theater
  - New Visualization Room in AERB coming in Spring 2021

NEW

UNIVERSITY of
HOUSTON
DIVISION OF RESEARCH

Carya: A Collaboration with Cullen College of Engineering

208 compute HPE nodes & 64 Nvidia Volta V100 GPUs (32Gb/e)

Theoretical peak performance of 770 Teraflops

10K CPU cores & 327K GPU cores

45 TB of main memory and 2 TB of high bandwidth GPU memory

Interconnect, Mellanox HDR Infiniband, 100Gb/s Line Rate

1,560 TB of shared hard-disk based storage

122 TB of shared flash storage

https://uh.edu/rcdc/resources/hpc/carya

UNIVERSITY of
HOUSTON
DIVISION OF RESEARCH

Sabine and Opuntia

Opuntia

80 compute nodes (64 GB/e) with 1,860 CPU cores
4 Nvidia K40 GPU nodes

3 large memory nodes (512-1000 Gb)

~600 TB shared storage
~1 TB/node of local compute-node space

Sabine

181 compute nodes with 5,704 CPU cores
8 nodes with 2 NVIDIA P100 GPUs each
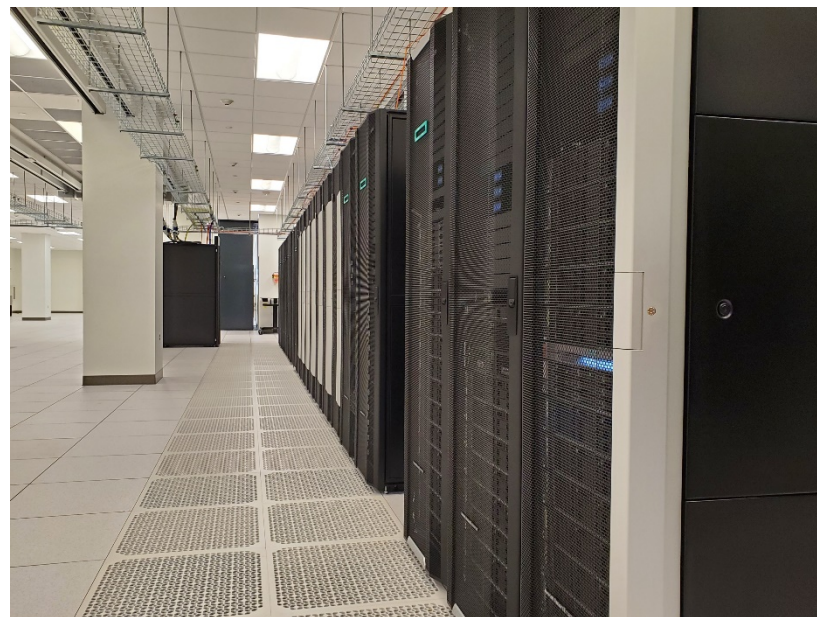4 nodes with 8 NVIDIA V100 GPUs each

128-256Gb of memory per node
1 HPE DL 360 large memory node 768Gb

~725TB NFS storage
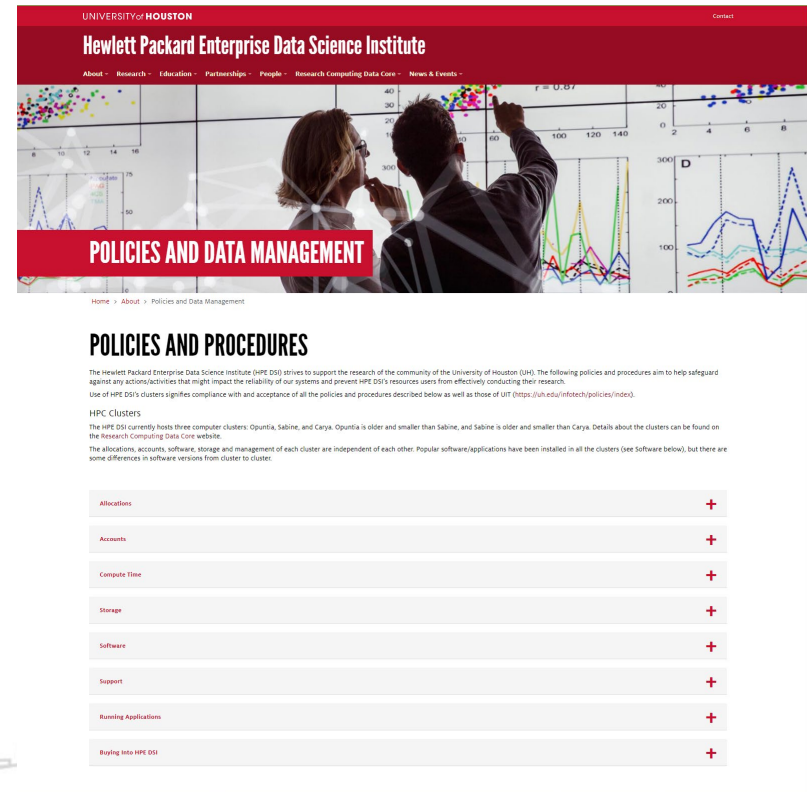
https://uh.edu/rcdc/resources/hpc

UH HPC Cluster Introduction Seminars

- Must participate before requesting an allocation

- On MS Teams

- ~20 participants/session

- For faculty, postdocs, and students

- Register at https://tinyurl.com/hpcintro

- 11/4 10-11:30
- 11/6 1-2:30
- 11/10 10-11:30
- 11/12 1-2:30
- 11/16 10-11:30

- 11/18 1-2:30
- 11/19 10-11:30
- More in Spring 2021

UNIVERSITY of
HOUSTON
DIVISION OF RESEARCH
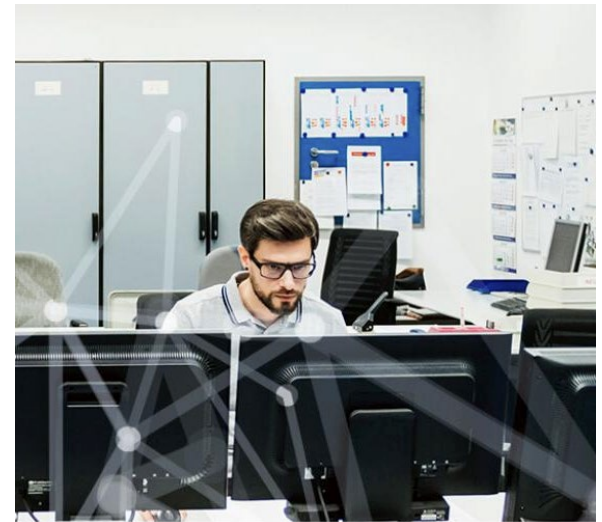
Storage and Allocations

- Yearly renewal cycle
  - September 1

- Use will determine cluster

- Option to buy nodes

- Storage limited
  - Shared resource
  - No archival storage
  - Option to buy storage

Allocations Size/Use Guidelines

	Opuntia [SUs, TB]	Sabine [SUs, TB]	Carya [SUs, TB]
Small	50k - 150k, 5	50k - 150k, 5	0, 0
Medium	250k - 999k, 5	250k - 999k, 10	0, 0
Large	0, 0	500k – 1.5M, 20	250k - 1M, 20
Huge	0, 0	1M – 3M, 20	1M – 3M, 20*

*The Allocations Committee may review time-limited proposals for special cases for 'Texas' projects which need more than 20Tb.

There goes Dr Pearcey and his personal computer

https://hpedsi.uh.edu/about/policies

UNIVERSITY of
HOUSTON
DIVISION OF RESEARCH

Courses

Jerry Ebalunode

Courses

- Core Courses
- 101 Introduction to Cluster Computing
- 201 Scientific Computing in C++
- 211 R for Data Science
- 212 Scientific Programming with Python
- 251 Data Visualization using Paraview and Tableau
- 301 Parallel Computing with OpenMP
- 302 Parallel Computing with MPI
- 303 GPU Parallel Programming
- 311 Introduction to Machine Learning

Data Science and HPC User Training Road Map

101 Intro to HPC

201 C++ | 211 R | 212 Python | 221 Julia

251 Visualization

301 OpenMP
302 MPI
303 GPU Programming

311 Machine Learning
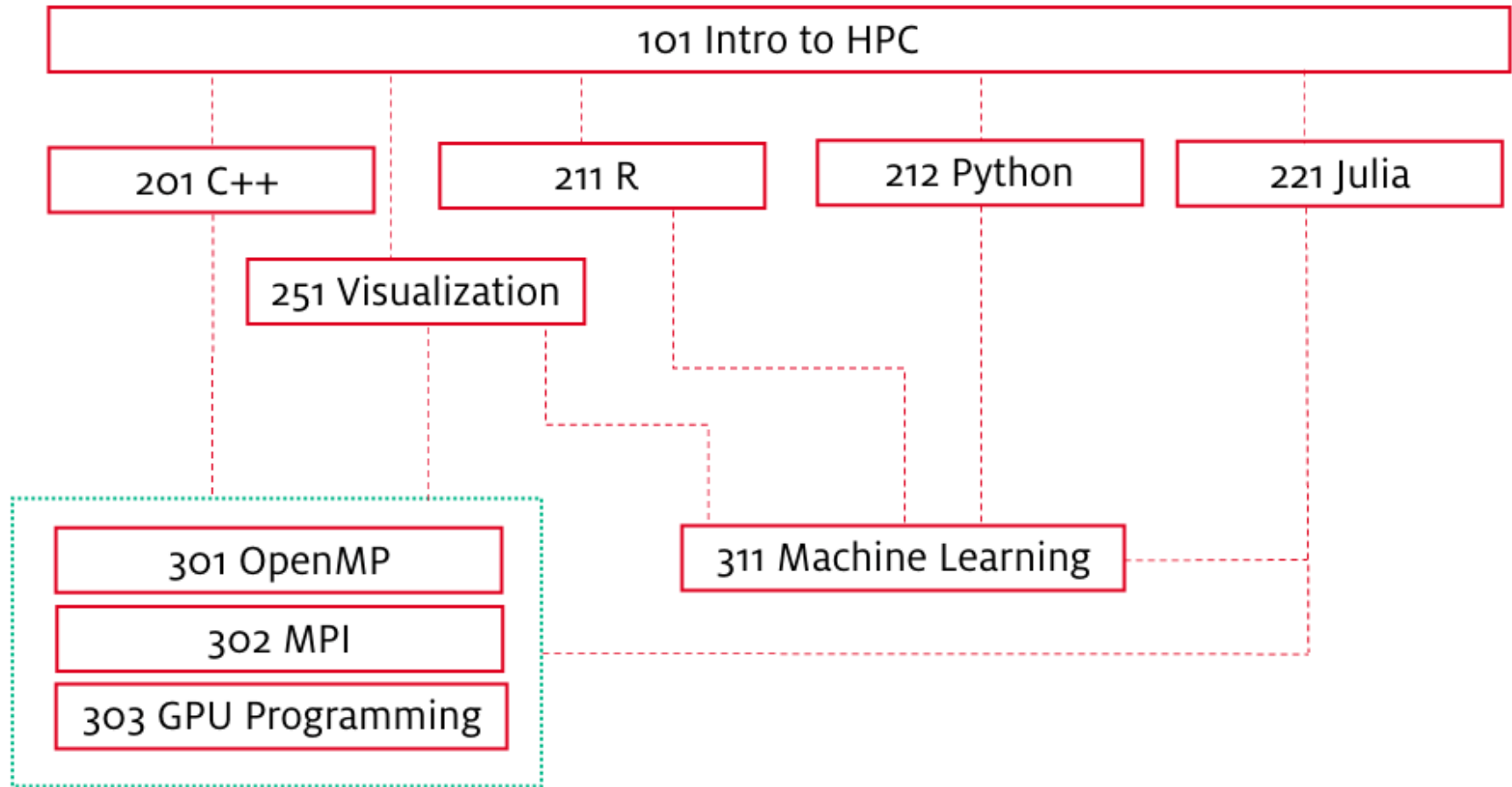
New Education Programs

Plans for Microcredentials

- Data Science Microcredential
    - 212 Scientific Programming with Python
    - 251 Data Visualization using Parview and Tableau
    - 311 Introduction to Machine Learning
- Application areas
    - Energy (partnership with UH Energy)
- Other areas in development

HPE Student Labs

- Funding available for undergraduate researchers to participate in collaboration between UH and external partners
  - Pilot projects over the summer
- Duration: 6 weeks to full semester
- Typically 10 hrs/week at $12/hr
- Eligibility: full-time student at UH
- Rolling deadline
- If interested, contact Claudia Neuhauser

UNIVERSITY of
HOUSTON
DIVISION OF RESEARCH

Engaged Data Science

Andrew Kapral

Engaged Data Science

- Summer programs
  - DASH
  - HPE DSI Girls Coding Camp
- PEERS
- Data and Society Minor
  - Collaboration with Honors College

Engaged Data Science - Summer Programs

Data Analytics in Student Hands (DASH) - Summer Research Fellowships

- **24 undergraduate students** – 10wk, full-time, online - $4K stipends
- **Partners:** Honors, Humana, CREATE, Conoco Phillips, UHCHWI
- **Projects:** 17 posters on display for [UH Undergraduate Research Day](#)
- **Students Profiles:** [Madhumitha Periyasamy](#) | [Sameer Sidiq](#)
- **Awards/Continuing Projects:** G. Litwin (DSI), S. Moursy (ARC), N. Siddeeque (ARC); E. Tran (ExCITE)

HPE DSI Girls Coding Camp

- **58 middle school students** – 80% girls
- **Partners:** College of Education – CodeCougars
- **Mentors:** Undergraduates paired with student teams
- **Topics:** HTML, Javascript, Python

Engaged Data Science – High School Mentoring

Project Engagement Encouraging Rising Students (PEERS)

- Local students mentored by UH undergraduates (near-peer)
  - **2018-20:** ~150 participants | **2020-21:** ~110 participants
- Project-based learning as a focal point for mentoring
  - Student teams develop projects in response to grand challenges 2019-20 - "What would you do with $5 million to improve preventative care in your community?"
- **Activities:** Weekly mentor meetings; workshops on college readiness; city-wide competition
- **2019-20 winners:** Lamar High School (teen health education)
- **APHA Presentation:** Undergraduate student leaders (HICH)

Engaged Data Science – Data & Society Minor

- Data and Society Minor
    - 15-hour program (open Honors) - launched fall 2020
    - **Pedagogy:** Project-based; community-engaged research
    - **Curriculum/Courses:** Interdisciplinary development; looking for elective courses across all disciplines.
    - **Approach:** Bring humanities questions to data science to create career pathways for people who need to communicate with data **Audience:** target people on management tracks to improve data-informed decision making
    - Hiring Visiting Assistant Professor (fall 2021) to expand instructional team.

Financial Sustainability

Funding

- No plans to charge for compute
- Reasonable limits on storage
- Possibility to purchase nodes (priority queue) and storage
- Funding sources
  - Base budget (DoR)
    - No IDC
  - Contracts and grants
  - Research funding for staff salary
  - Certificate programs and microcredentials
  - Gifts
    - HPE Donation
  - …

UNIVERSITY of
HOUSTON
DIVISION OF RESEARCH