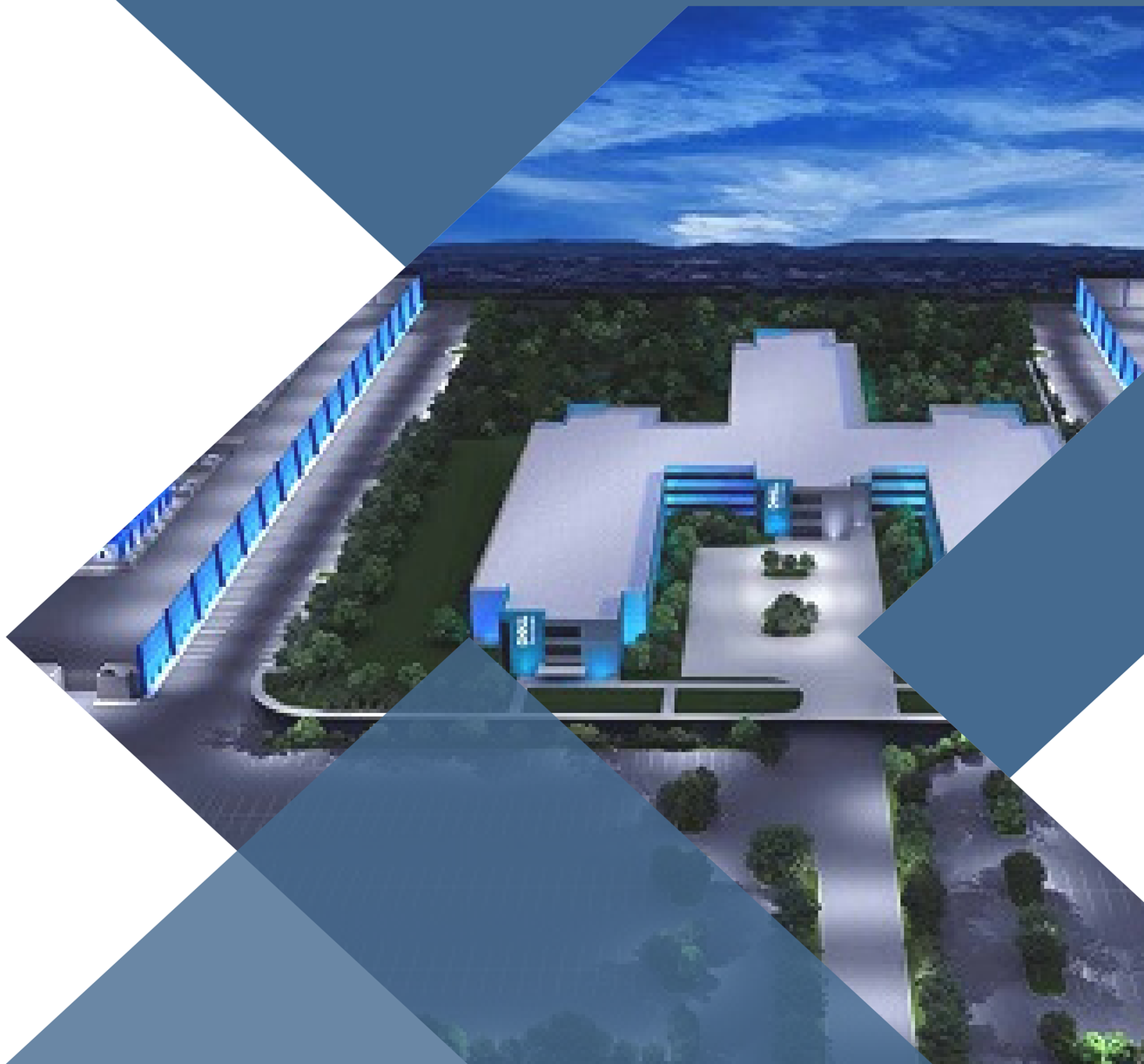
Texas Advanced Computing Center

TACC



Our Team Members

Raja Allmdar Tariq Ali

Indiana University - Purdue University, Indianapolis, 2026
Computer Science B.S.

Natalie Mason

Arizona State University, 2024
Information Technology B.S.

Ellie Fahey

University of Notre Dame, 2025
Applied Computational Math and Statistics B.S.

Saad

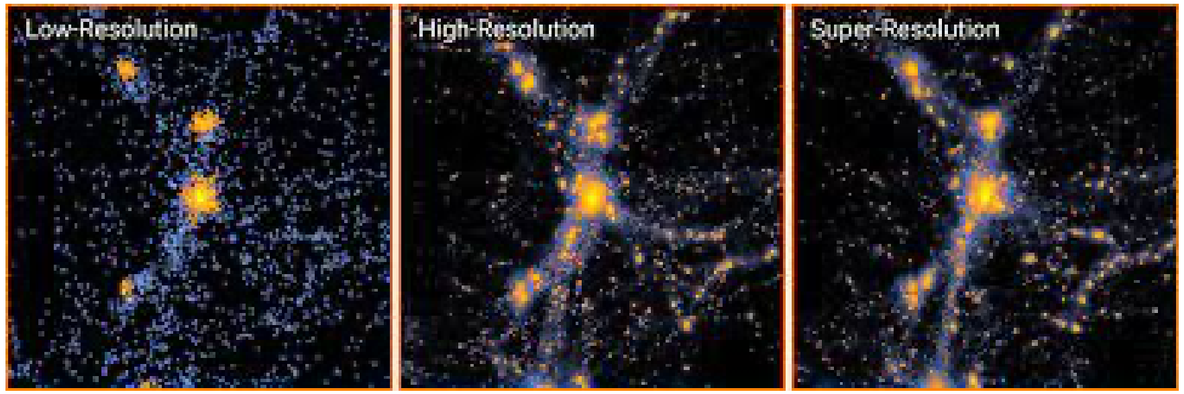
University of Alabama at Birmingham, 2026

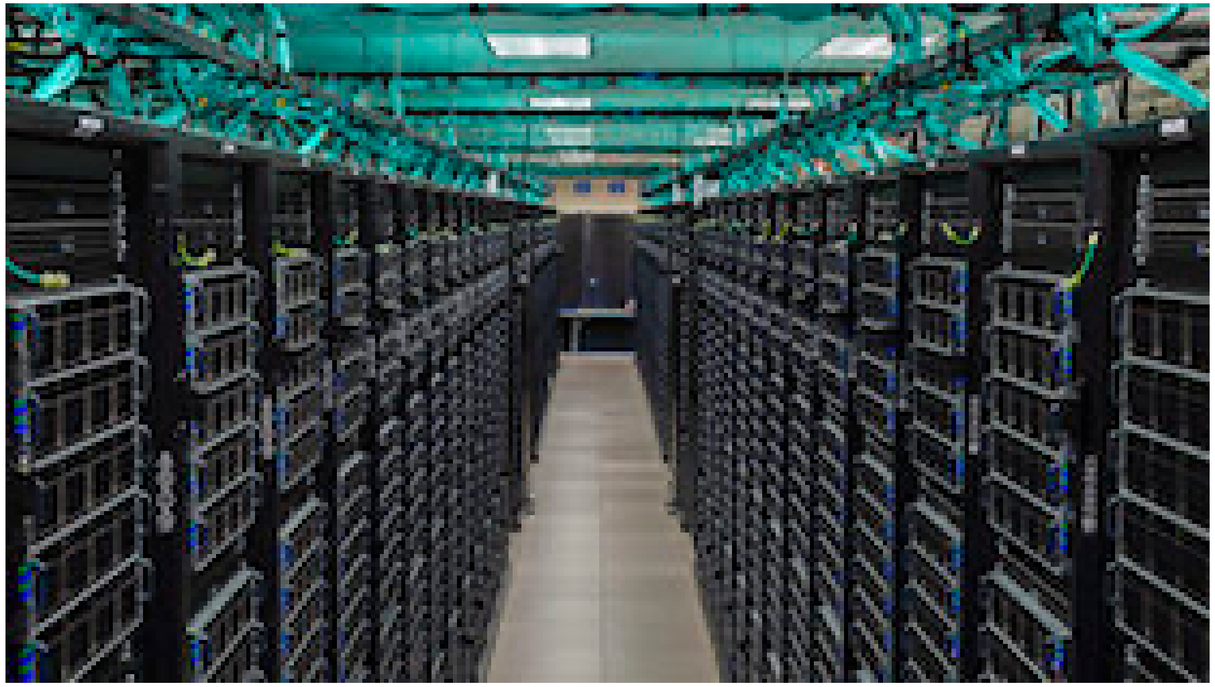
What is TACC ?

1. Center of Computational Excellence
2. Advanced Computing Technologies
3. Federally Funded Projects

Research Contributions

How does TACC contribute to advancing scientific research and discovery across multiple disciplines?

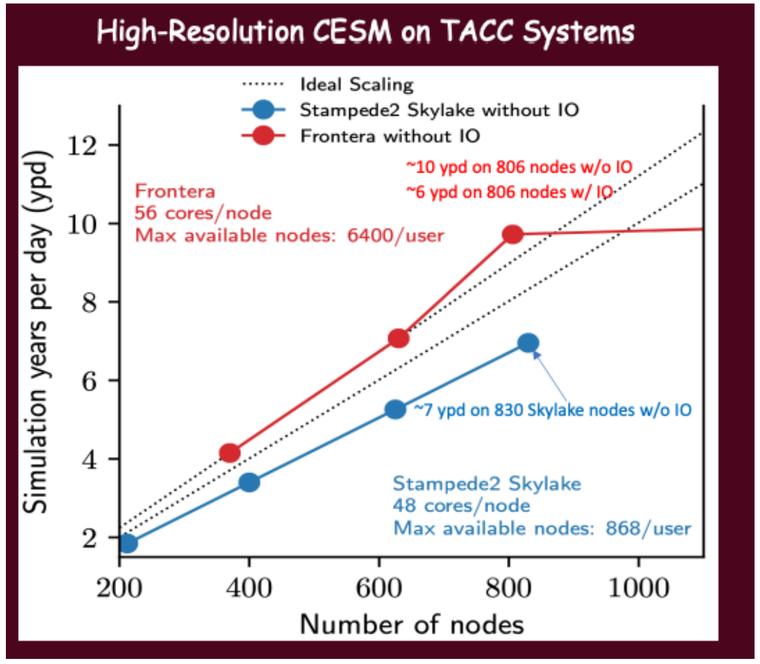


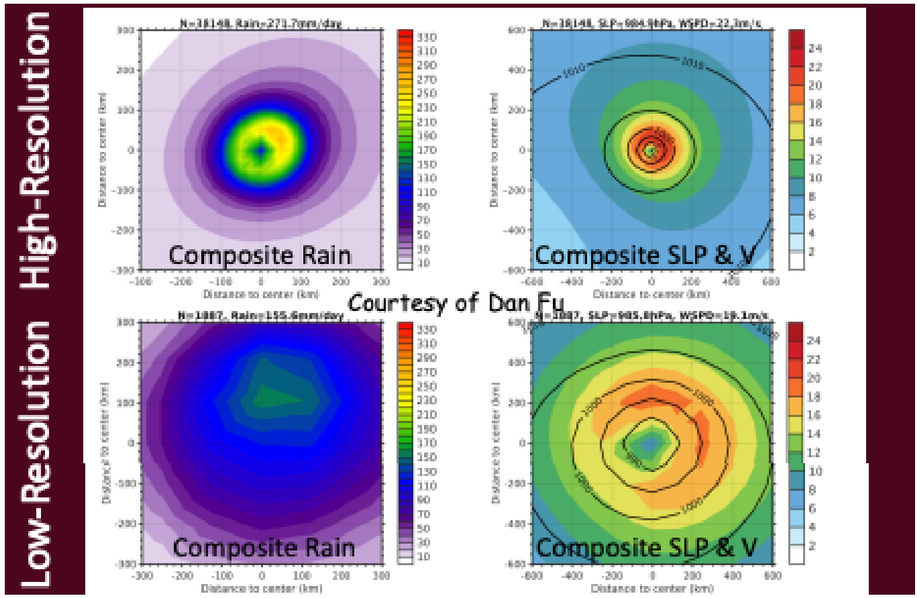


TACC is a top research center offering interdisciplinary collaborations, and support services for diverse range of research

Climate Modeling and Weather Prediction:

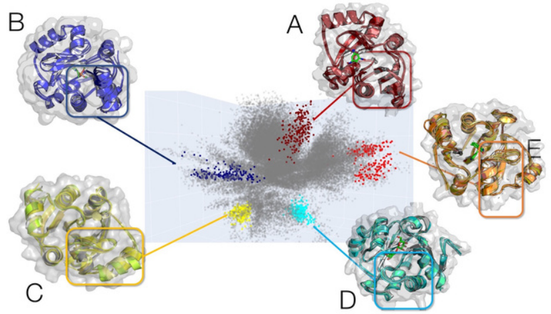
- A high-resolution climate model, the Community Earth System Model (CESM) project with NCAR utilized TACC's resources (Frontera SuperComputer) to simulate global climate patterns over decades, providing valuable insights into the consequences of increasing greenhouse gas emissions.





Drug Discovery and Bioinformatics:

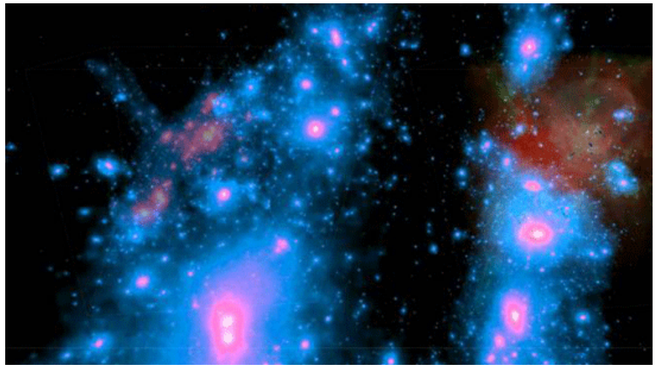
- TACC's computing resources expedite drug discovery by allowing rapid screening of numerous compounds, with programs like *DrugDiscovery@TACC* identifying promising candidates for diseases like cancer and Alzheimer's through virtual screenings and molecular simulations.



TACC using drug discovery to fight COVID-19

Astrophysics and Cosmology:

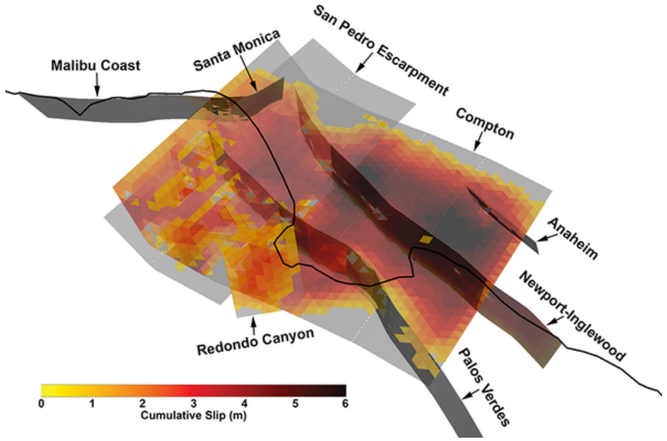
- The *Illustris Project* utilized TACC's Stampede and Ranger supercomputers for highly detailed universe simulations, enhancing our knowledge of galaxy formation and evolution.



Simulation produced a model of the universe's evolution

Earthquake modeling and hazard assessment:

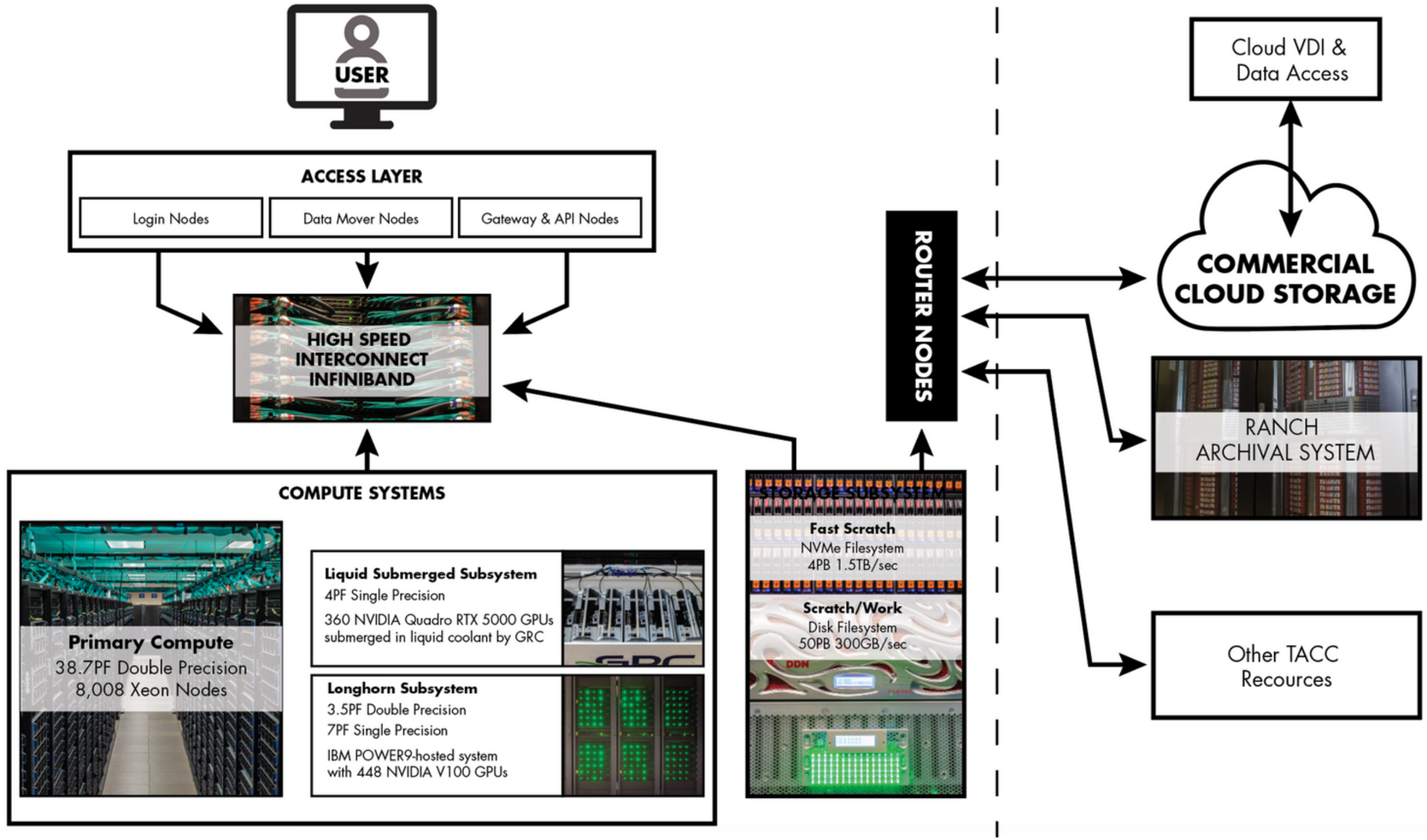
- The *Southern California Earthquake Center (SCEC)* used TACC's supercomputers, such as Frontera, to simulate hundreds of thousands of years of seismic history in California, generating realistic earthquake probabilities and improving hazard assessment.



3D view of one especially complex multi-fault rupture from the synthetic earthquake catalog

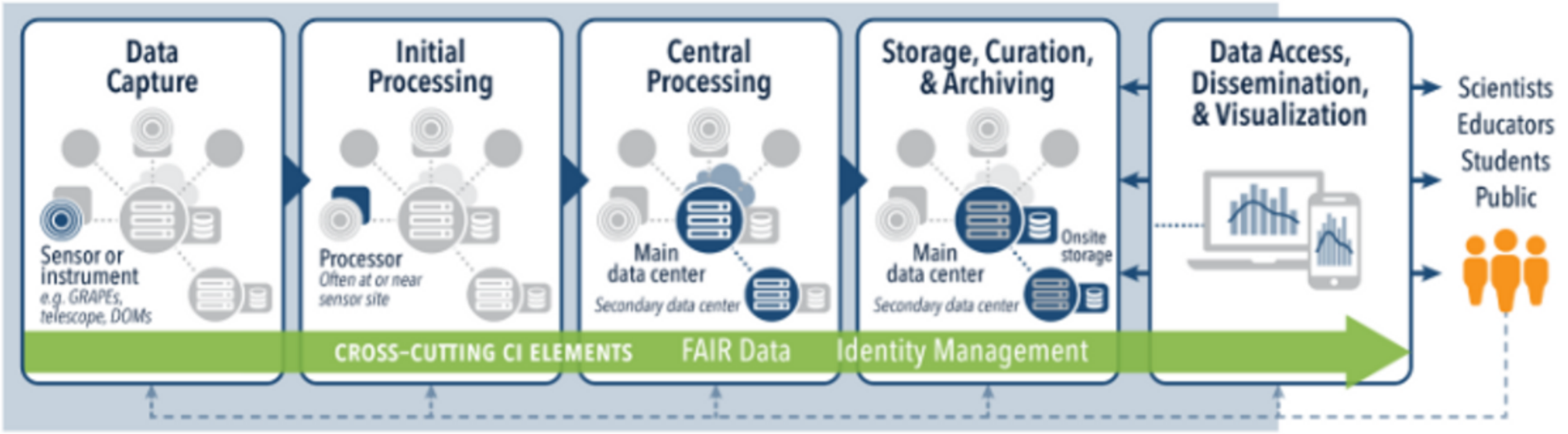
Frontera

Supercomputing at TACC



Data life cycle

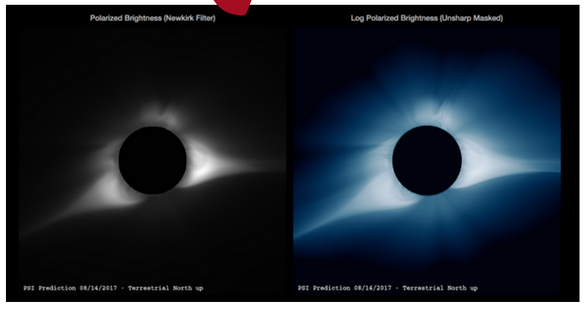
How is TACC managing its data life cycle, from data capture and processing to archival dissemination, and what tools and technologies are being used to support these activities?



What type of Data is TACC creating or collecting?

Simulation Data

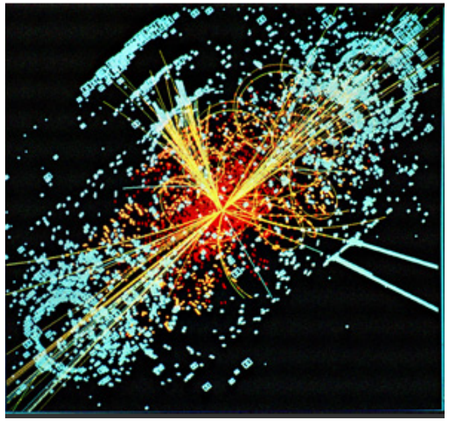
- Refers to output of computer models and simulations that represent real-world processes, systems or phenomena.
- Data helps scientists better understand complex phenomena, predict outcomes and optimize processes.
- Some examples include molecular dynamics, climate, fluid dynamics, and astrophysical simulations



researchers used computer simulations to predict the appearance of a solar eclipse, specifically the solar corona's structure during the eclipse.

Experimental Data

- Refers to the information collected from controlled experiments or observations conducted by researchers in various scientific domains.
- Researchers generate experimental data through a wide range of techniques and methods, including genomics, High-Energy Physics, and Neuroscience Data.



a project where researchers use TACC's computational resources to study Technicolor theory, an extension of the Standard Model of particle physics that proposes the existence of new particles and forces.

Observational Data

- Refers to information collected from real world situations.
- Direct measurements, monitoring, recording
- Variety of scientific domains (astronomy, geophysics, medicine)
- TACC storage systems (Corral, Ranch) for effective data management and security
- Supports data sharing, collaboration by data publication services for community

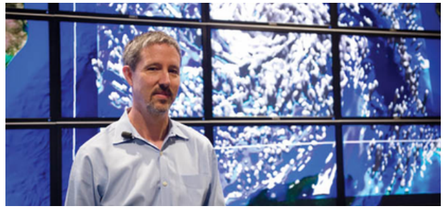
DESIGNSAFE-CI
A NATURAL HAZARDS ENGINEERING COMMUNITY

A Cyberinfrastructure project that provides tools for researchers studying natural hazards. Observational data involves information from sensors, monitoring stations, and remote sensing technologies like satellite imagery, used to study natural hazards' impacts and develop resilient structures and systems.

Data Products generated from Raw Data

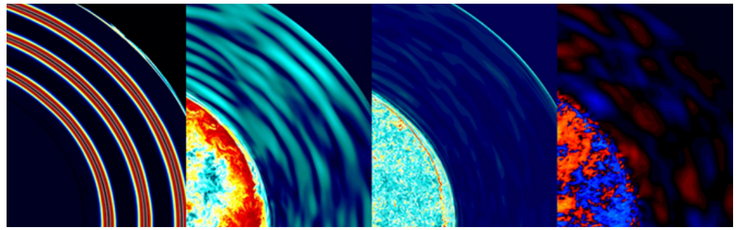
Processed Data

- Raw data cleaned, pre-processed, and transformed for analysis
- Prof. Dawson's team studies storm surge and coastal flooding
- ADCIRC model processes raw data to create high-resolution storm surge predictions and flood inundation maps for improved hurricane preparedness and coastal resilience



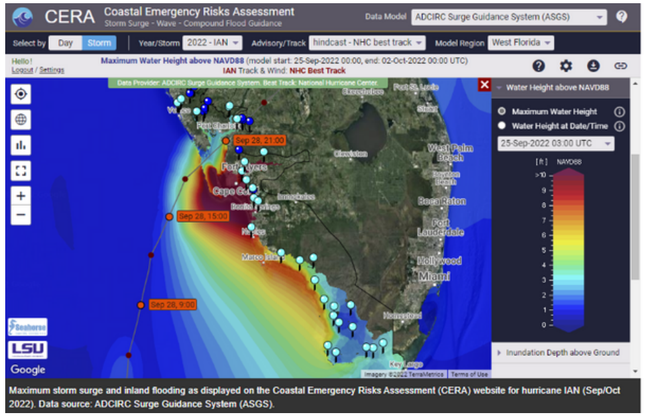
Statistical Analysis

- Researchers perform statistical analyses on data to identify trends.
- Researchers at UT Austin have used TACC's Stampede supercomputer to analyze gravitational wave data from the LIGO and Virgo observatories.
- Processed raw data from gravitational wave detectors
- Applied statistical analysis to determine the significance of these signals and distinguish them from background noise.



Visualization

- Researchers may create visualizations of their data to better understand patterns, relationships, and trends.
- Carola Kaiser and her team at LSU CCT utilized TACC's Lonestar6 supercomputing system to run the ADCIRC model for Hurricane Ian simulations



Maximum storm surge and inland flooding as displayed on the Coastal Emergency Risks Assessment (CERA) website for hurricane IAN (Sep/Oct 2022). Data source: ADCIRC Surge Guidance System (ASGS).

Data Capture

- Data capture refers to the process of collecting and ingesting raw data from various sources such as sensors, monitoring stations, experiments, simulations, and remote sensing technologies.
- TACC facilitates data capture by offering access to several tools designed to handle large-scale data collection such as:

High-Performance Computing (HPC) systems:

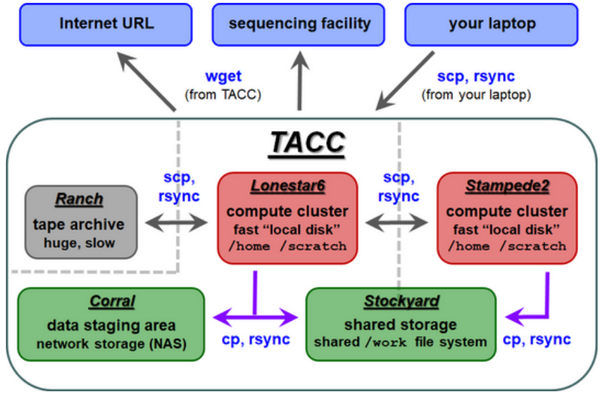
- TACC's HPC systems, such as **Frontera, Stampede2, and Lonestar5**, facilitate **data generation** and capture by supporting large-scale data handling and complex calculations in simulations, experiments, and observations.

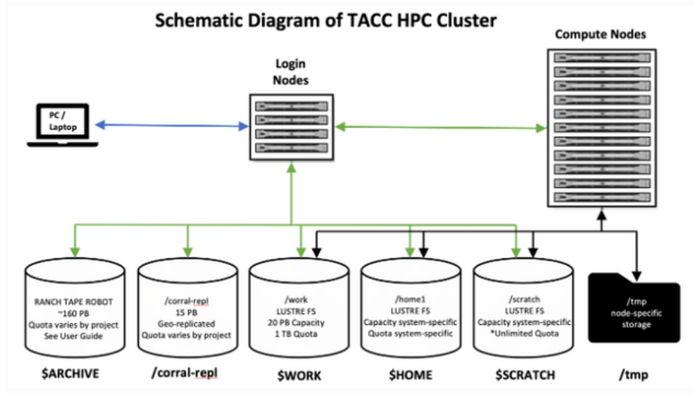
Data Storage systems:

- TACC provides researchers with storage systems like Ranch (mass storage), Corral (large-scale data collections), and Stockyard (global file system), enabling efficient **data ingestion**, storage, organization, and management of raw data.

Data Sharing and Research Platforms:

- TACC's **DesignSafe** platform enables data capture, management, and sharing for natural hazard research while promoting accessibility and collaboration through services like TACC's Data Collection Catalog.

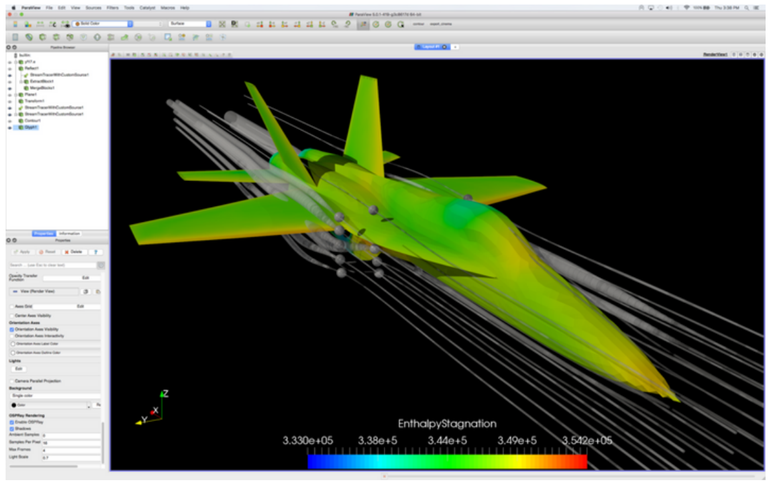




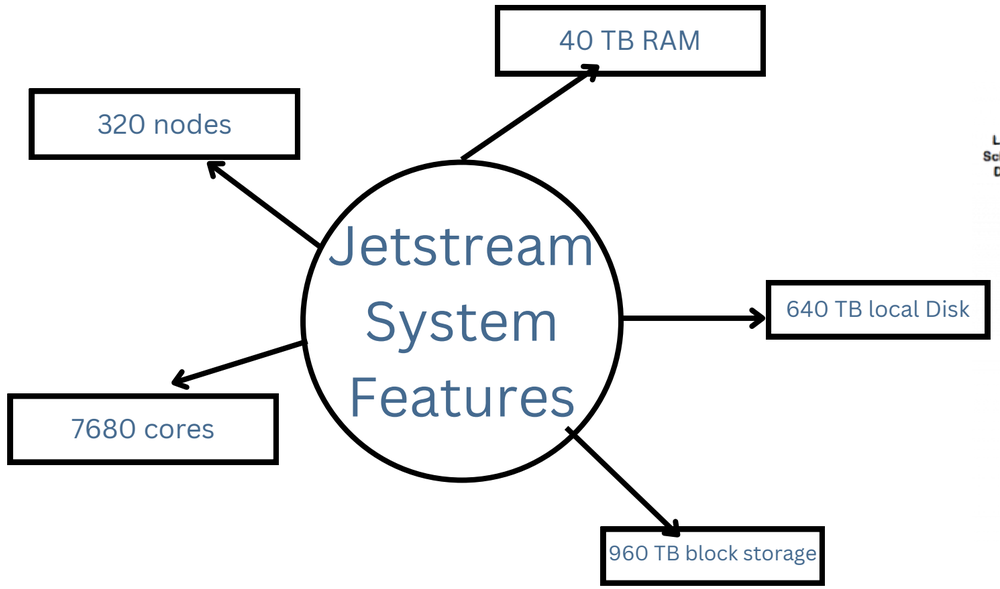
Data Processing

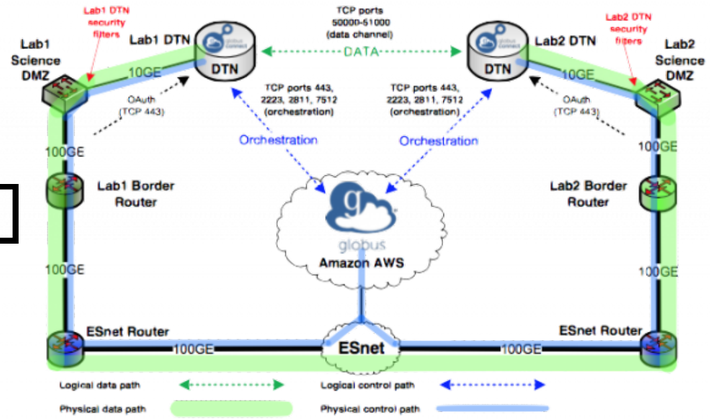
Software and Tools:

- TACC offers several visualization software packages, including **VisIt** and **ParaView**, to help researchers create visual representations of their data.
- TACC utilizes **MPI** and **OpenMP**, which are popular parallel programming models that enable efficient processing of large datasets in high-performance computing.
- TACC's **Jetstream** platform provides pre-configured virtual machines with pre-installed software stacks, enabling researchers to efficiently access specialized environments for data processing, analysis, modeling, visualization, and machine learning without manual installation and configuration.
- **Globus** and **Data Transfer Node (DTN)** are essential tools for data capture and processing. Globus simplifies the process of **transferring** large datasets between different storage systems, while DTN provides high-speed data transfer capabilities.



TACC collaborates with Kitware and Intel to integrate OSPRay ray-tracing capabilities in visualization tools like ParaView and VisIt, enhancing data processing and providing high-fidelity rendering for complex data in high-performance computing platforms

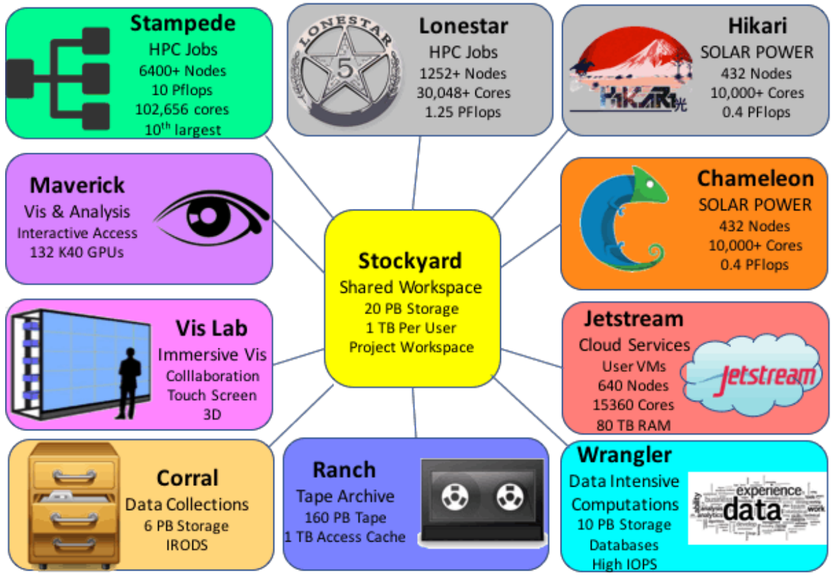




Data Storage

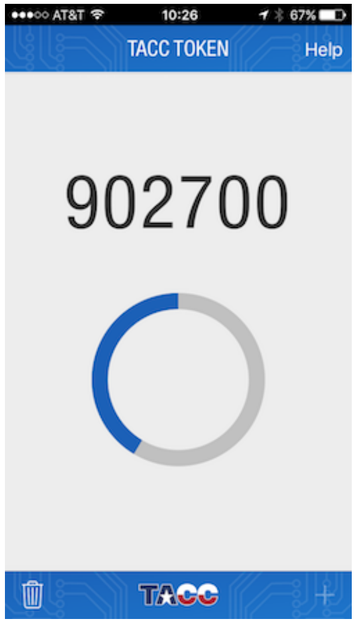
Data Management Tools:

- TACC uses **iRODS**, an integrated data grid and data management tool, to store data in a unified namespace, replicate data across multiple systems, and store checksums and metadata for files, enabling efficient data management and collaboration for researchers.



Security and Access Controls:

- TACC developed a cost-effective and flexible authentication solution called OpenMFA by integrating LinOTP, RADIUS, HTTPS, and custom pluggable authentication modules (PAM) to securely enable tens of thousands of researchers to access its high-performance computing systems while maintaining ease of use.

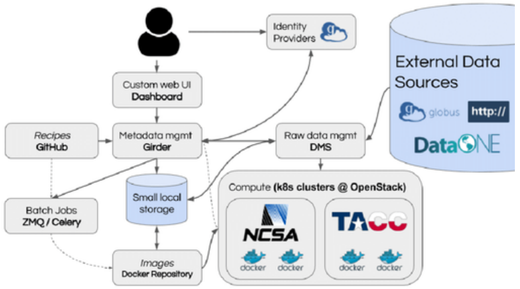


- TACC ensures data storage security and access controls by employing Access Control Lists (ACLs), which are powerful tools that allow for specific users and groups to be granted read, write, and/or execute permissions on any file or directory, providing a more flexible mechanism than the traditional Unix permissions system.
- TACC employs encryption algorithms like **AES-256-CBC** for secure data transfers, minimizing the risk of data breaches during the transfer process. This ensures the confidentiality and integrity of data as it moves between systems.

```
"irods_encryption_key_size": 32,
"irods_encryption_salt_size": 8,
"irods_encryption_num_hash_rounds": 16,
"irods_encryption_algorithm": "AES-256-CBC",
"irods_default_hash_scheme": "SHA256",
```

DESIGNSAFE-CI
A NATURAL HAZARDS ENGINEERING COMMUNITY

What are the best practices for data management, including data cleaning, preprocessing, storage, and sharing, that are being used at TACC, and how can these best practices be transferred to other scientific research domains?



Data Storage:

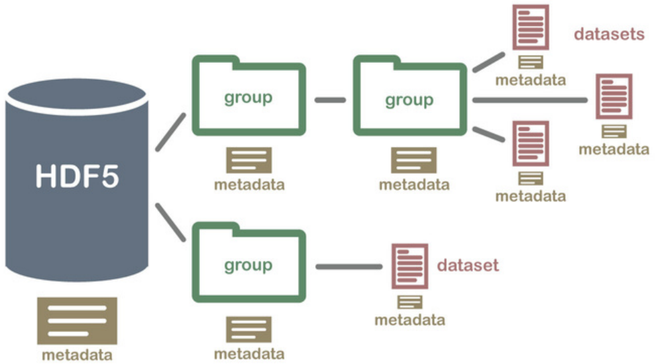
- Standardize formats and conventions, like dates and units, for consistency across datasets.
- Choose the right storage solution based on the data's size, type, and access requirements. TACC offers various storage systems, such as Ranch (long-term archival storage), Corral (project storage), and Stockyard (high-performance parallel file system).
- Employ version control systems to track changes to the data and facilitate collaboration among researchers.

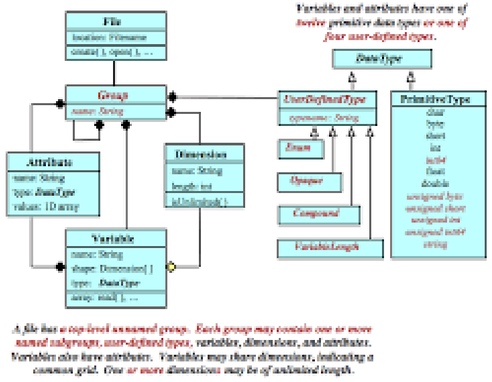
Data Sharing:

- Utilize platforms like TACC Science Gateway or iRODS (Integrated Rule-Oriented Data System) to facilitate data sharing, collaboration, and discovery among researchers.
- Implement proper access controls, such as Access Control Lists (ACLs), to manage access permissions for different users or groups, and ensure data security.

Documentation, Metadata and adherence to standards :

- Document data collection methods, processing steps, and analysis techniques to promote transparency and reproducibility in research.
- Create comprehensive metadata that describes the data, its origin, and any associated data processing steps. This will help others understand and use the data effectively.
- Employ standardized data formats, such as NetCDF or HDF, that are widely recognized and supported in the scientific community.





TACC

THANK YOU!

Website

Kahoot!

Please feel free to ask questions about the topics we discussed today!