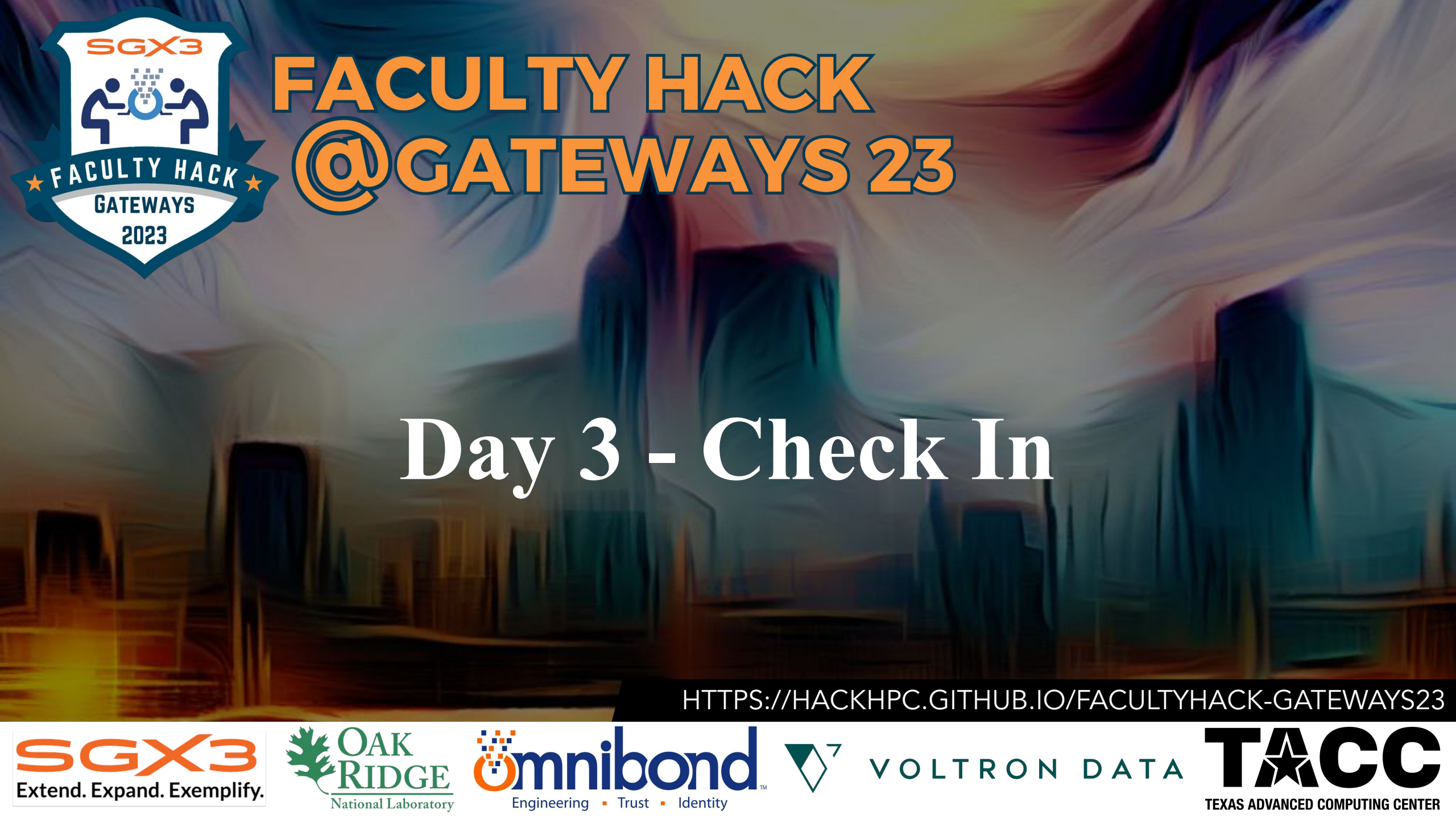
SGX3
FACULTY HACK
GATEWAYS
2023
FACULTY HACK
@GATEWAYS 23
Day 3 - Check In
HTTPS://HACKHPC.GITHUB.IO/FACULTYHACK-GATEWAYS23
SGX3
Extend. Expand. Exemplify.
OAK RIDGE
National Laboratory
Omnibond
Engineering ▪ Trust ▪ Identity
VOLTRON DATA
TACC
TEXAS ADVANCED COMPUTING CENTER

Instructions

1. 3 Slides, 3 minutes
2. Add your team information to the slide
   a. Course Description
   b. Resources
   c. Sample Datasets (if applicable)

Practical Introduction to HPC and Research Computing

CAMSA Team

CAMSA
FACULTY-HACK

1. Course Description

- This course is meant to students but also to faculty/staff beyond typical course environments (CSCI 5306)
- The main goal is to create a course that can be flexible in the time/duration from 1 to 16 weeks (1. Adding more training and modules, and 2: Produce customized modules to the different disciplines).
- The course will consist of lectures, practical hands-on homework assignments, and hands-on laboratory work where students will try to build their own scripts to be executed in the HPC environment.

2. Potential resources/pre-reqs

- Resources: HPC websites of public institutions, examples:
  - https://access-ci.org/
  - https://hprc.tamu.edu/
- Pre-requisites:

- Ideally, I want pre-reqs to be minimum so that such course can be open to all students and faculties from all disciplines,

3. Identifying Sample Datasets

- https://towardsdatascience.com/ (requires an account)
- Kaggle datasets and codes
- Github repositories
- https://medium.com/
- But those are very broad, I couldn’t see things that are more specific, specially per domain of knowledge

Team Name: ThreatTracker

Computing Tools/Environment

- GitHub (to store code and data) (optional)
- Python 3.8+ with packages (faker)
- Oasis stix2-generator, stix2-validator, stix-visualizer
- Synthetic Data Vault
- MITRE ATT&CK STIX Data

Skills/Knowledge/Abilities

- Python
- Statistics
- Databases
- Basic cyber intrusion knowledge

Course Assessment

- 25% of the overall grade: Create frontend for Identity, Malware, and Threat Actor objects
- 25% of the overall grade: Generate STIX objects from user input, Finish STIX objects and store them in the database
- 25% of the overall grade: Generate/visualize a graph using three STIX objects Identity, Malware, and Threat Actor
- 25% of the overall grade: Anomaly detection using Deep Learning Algorithms.

Theme Song:

https://soundcloud.com/alsyn/synesthesia?in=sc-playlists/sets/brainwaves

UNC PEMBROKE

Threat Tracker Team

- **Target course:** CYB 4900 Cybersecurity Capstone Project
- **Course Description:** In this course, student integrates deep learning with cybersecurity threat intelligence to address the specific challenges posed by Internet-of-Vehicles, particularly in the context of emergency vehicles using synthetic cyber knowledge graphs to represent and analyze cyber threat intelligence and relationships, and they will employ deep learning algorithms such as Autoencoders, RNNs, and CNNs for anomaly detection within this graph data.
- **Potential Resources (Tools,Packages,IDEs)**:

MITRE ATT&CK, Synthetic Data Vault, Python 3.8+ (faker, pandas, etc libs), Oasis STIX2 (generator, validator, visualizer), Repos-IDEs-HPCs (Omnibond, GitHub, Jupyter, Sagemaker CPUs, ArgonneLabs GPUs, OakRidgeLabs GPUs)

- **Dataset Resources**: Kaggle, UCI ML Repo, Google Datasets

Team Altair

Bernie Boscoe, Southern Oregon University

Team Mentors : Veronica Vergara & Mohamed Elbakary

Team Theme Song: New Order, Thieves like us remix (1987)
https://soundcloud.com/markaymufc/new-order-thieves-like-us-mk-instrumental-cover-kleptomaniac-mix

Goals:

To add a module to an undergraduate Intro to Data Science course that demonstrates how to use Jupyter Notebooks in the cloud, with a large dataset, and if I can, GPUs to train an ML model that would not be possible to do without a GPU-enabled device. Outcomes would be an understanding of accessing cloud interfaces, basic terminal commands, an overview of the Jupyter notebook as both a local and cloud tool, and if possible, how to test if GPUs are being seen. Update: possibly using JetStream2

What I need help with: what resources have Jupyter notebooks with GPU option? How can we all share a space, for example for 25 students? How do I handle accounts? How can we load/make available a dataset for them to access?

https://github.com/bboscoe/gateways23

Jarvis Bulldog Team

Team Members: Widodo Samyono,
Jarvis Christian University

Team Mentors: Je'aime Powell
TACC

Team Theme Song

i. Song name : Hey Bulldog
ii. Artist : The Beatles
iii. URL Link to the song: https://www.youtube.com/watch?v=M4vbJQ-MrKo

Jarvis Bulldog Team

Our Goals:

1) Redesigning MATH 3390: Computational and Mathematical Biology, using HPC Open Sources from Science Gateways.
2) Building a website for MATH 3390: Computational and Mathematical Biology, using the HPC Open Sources.
3) Piloting the redesigned course in Spring 2024.
4) Conducting surveys and evaluations for the course.

Url to our team GitHub repository:
https://github.com/wsamyono/BulldogTeamFacHackGA23

Jarvis Bulldog Team

MATH 3390: Computational and Mathematical Biology

Course Description

In some cases, it is too dangerous or impossible to do an experiment, so we can do numerical experiments through mathematical modeling and simulation. Besides learning mathematical modeling, the students in this course will learn basic commands, syntaxes, and fundamental programming in Python and use them for solving problems in biology. The course targeted students having major in mathematics, and biology and chemistry with minor in mathematics who are interested in learning computational and mathematical biology. The course consists of 3 parts: 1) fundamental programming in Python, 2) computational biology, and 3) mathematical biology.

Jarvis Bulldog Team

Potential Resource Needs

1) Google Colab
2) Yupiter Notebook
3) Anaconda Navigator
4) Python
5) SciPy
6) Sklearn
7) Others from Science Gateways including TACC and ACCESS

Note: I have an account for TACC: Texas Advanced Computing Center (utexas.edu)

Jarvis Bulldog Team

Sample Datasets

1) RCSB PDB Protein Data Bank: https://www.rcsb.org/
2) Genomic Data Commons Data Portal: https://portal.gdc.cancer.gov/
3) Data from students in Biology conducting in vitro experiments by inducing nanoparticles into cancer cells. Data can be acquired directly from the students and the biology faculty members.
4) Other biology data from Science Gateways.

Team Tech Tigers

Team Theme Song

i. Song name : Weird Science
ii. Artist : Oingo Boingo
iii. URL Link to the song: https://soundcloud.com/oingo-boingo-official/weird-science-album-version?si=e08c2d1f6ce54be18aa649d1ea08556c&utm_source=clipboard&utm_medium=text&utm_campaign=social_sharing

Bulldogs Team

Team Member: Dr. Rui Zhu
(Kettering University)

Mentor: Dr. John Holmen
(Oak Ridge National Laboratory)

Mentor: Yvonne Phillips
(Morehouse College)

- Target Course(s): CS425 Parallel Programming and Algorithms, CS457 Wireless and Mobile Security
- Goal:
  - Integrating HPC with Cybersecurity, Cryptography, and Machine Learning to develop curriculums
  - Identify applicable HPC resources from ORNL/wider HPC community and develop course descriptions
  - Create and refine course schedules, hands-on labs, etc.
- GitHub Repo: https://github.com/ruikobe/KetteringTeamFacHack23
- Theme Song: George Thorogood & The Destroyers - Bad To The Bone

Course Description

- The CS-425 **Parallel Programming and Algorithms** course introduces students to the foundations of parallel computing.

- The course will include material on emerging multicore hardware, shared-memory programming models, message passing programming models used for cluster computing, data-parallel programming models for GPUs, and problem-solving on large-scale clusters using MapReduce.

- A key aim of the course is for students to gain a hands-on knowledge of the fundamentals of parallel programming by writing efficient parallel programs using some of the programming models that students learn in class.

Topics

1. Introduction to Parallel Computing
2. Parallel Programming Platforms
3. Principles of Parallel Algorithm Design
4. Basic Communication Operations
5. Analytical Modeling of Parallel Programs
6. Programming Using the Message Passing Paradigm, e.g., Message-Passing Interface (MPI)
7. Programming Shared Address Space Platforms
8. Dense Matrix, Sorting, Searching, and Graph Algorithms
9. Graphics Processing Units (GPUs)
10. Compute Unified Device Architecture (CUDA)

Potential HPC Resources

- A few courses bringing together parallel programming, parallel algorithm, and HPC:

  https://www.cs.purdue.edu/homes/ayg/CS525/index.html

  https://faculty.cc.gatech.edu/~umit/GT/CSE/2020/CSE6230.html

- Training archives from some of the larger HPC centers:

  https://www.alcf.anl.gov/support-center/training-assets

  https://docs.alcf.anl.gov/account-project-management/allocation-management/overview/

  Allocations at various HPC center:

  https://docs.alcf.anl.gov/account-project-management/allocation-management/overview/

  https://www.chpc.utah.edu/userservices/allocations.php

  https://docs.olcf.ornl.gov/accounts/accounts_and_projects.html

  https://tacc.utexas.edu/use-tacc/allocations/

Sample Datasets

- CRAWDAD dataset

  There are huge data sets in different fields, e.g., cybersecurity, wireless networking, IoT, Transportation, Power and Energy, etc.

- BLE-WBAN: RF real-world dataset of BLE devices in human-centric healthcare environments

  In communication and networking research, obtaining large, real-world datasets related to the physical layer has always been challenging, especially in IoT and Health IoT.

- Dataset from HPC center of ORNL, and other HPC centers

@MacCarthy@Kamanpiu@Alsmadi

WorkFlow Labs in HPC
CAMSA Team

Something that always work (Reproducibility)

CAMSA
FACULTY-HACK

Goals:
1. Integrating HPC within CSCI 5306 (Computer Networks) Workflows lab
2. Produce introductory material to all other majors

Ballad of the Alamo, R.W. Hampton, https://soundcloud.com/r-w-hampton/ballad-of-the-alamo?in=user-470687170/sets/normal

https://github.com/alsmadi/CAMSA_Gateways_2023

FACULTY HACK@GATEWAYS 23

HTTPS://HACKHPC.GITHUB.IO/FACULTYHACK-GATEWAYS23

Supported by:

SGX3
Extend. Expand. Exemplify.

OAK RIDGE
National Laboratory

Omnibond
Engineering ▪ Trust ▪ Identity

VOLTRON DATA

TACC
TEXAS ADVANCED COMPUTING CENTER

HACKHPC.ORG