Interactive Large-Scale Data and Graph Analytics

Introduction

Oliver Alvarado Rodriguez, Naren Khatwani, **Zhihui Du (Presenter)**, David Bader

NJIT New Jersey Institute of Technology

Arachne

A graph infrastructure enabling more
efficient data science


NJIT
New Jersey Institute
of Technology

New Jersey Institute of Technology

"NJIT Climbs the Rankings of U.S. News & World Report, A Top 50 Public University" – 13 Sep 2021

"NJIT Named As One of Nation's 'Best Colleges' for 2022, The Princeton Review Says" – 6 Sep 2021

NJIT
New Jersey Institute of Technology

Launched in **July 2019**, with inaugural director
David A. Bader
(~40 faculty in current centers)

Center for Big Data	• Big Data Analytics, Systems, and Tools • Cyberinfrastructure
Cybersecurity Research Center	• Practical encryption • Privacy technologies • Information Assurance
The Structural Analysis of Biomedical Ontologies Center	• Medical Informatics • NIH / National Cancer Institute
FinTech Group	• Financial Services • Insurance Industry
Machine Learning & AI	• Real-world technologies • Industrial partnerships

Solving real-world challenges

- Urban sustainability
- Healthcare analytics
- Trustworthy, Free and Fair Elections
- Insider threat detection
- Utility infrastructure protection
- Cyberattack defense
- Disease outbreak and epidemic monitoring

NJIT
New Jersey Institute of Technology

NJIT New Jersey Institute of Technology

Alumni Giving MyNJIT Login Tools Search NJIT

Admissions Academics Research About Life at NJIT Athletics I AM A... APPLY NOW

Institute for Data Science Aims to Democratize Supercomputing With NSF Grant

Written by: Evan Koblentz

Published: Wednesday, March 17, 2021

New algorithms from at NJIT can make supercomputer power available to almost anyone

Ordinary people could soon have greater ability to analyze massive amounts of information, based on new algorithms and software tools being designed at NJIT, intended to simplify

https://news.njit.edu/institute-data-science-aims-democratize-supercomputing-nsf-grant

High Performance Algorithms for Interactive Data Science at Scale

(PI: Bader)
3/2021 – 2/2023
NSF CCF-2109988

NSF

αρκούδα massive scale data science

NJIT New Jersey Institute of Technology

Publications

- **Oliver Alvarado Rodriguez, Zhihui Du, Joseph Patchett, Fuhuan Li, David Bader (2022). Arachne: An Arkouda Package for Large-Scale Graph Analytics. IEEE HPEC.**
- Zhihui Du, Joseph Patchett, Oliver Alvarado Rodriguez, Fuhuan Li, David Bader (2022). High-Performance Truss Analysis in Arkouda. IEEE HiPC.
- Joseph Patchett, Zhihui Du, Fuhuan Li, David Bader (2022). Triangle Centrality in Arkouda. IEEE HPEC.
- Zhihui Du, Oliver Alvarado Rodriguez, David Bader (2021). Large Scale String Analytics In Arkouda. IEEE HPEC.
- Zhihui Du, Oliver Alvarado Rodriguez, David Bader (2021). Enabling Exploratory Large Scale Graph Analytics through Arkouda. IEEE HPEC.
- Joseph Patchett, Zhihui Du, David Bader (2021). K-Truss Implementation in Arkouda (Extended Abstract). IEEE HPEC.
- Zhihui Du, Oliver Alvarado Rodriguez, Joseph Patchett, David Bader (2021). Interactive Graph Stream Analytics in Arkouda. Algorithms.
- Zhihui Du, Oliver Alvarado Rodriguez, David A. Bader, Michael Merrill, William Reus (2021). Exploratory Large Scale Graph Analytics in Arkouda. CHIUW.

NJIT
New Jersey Institute of Technology

Arkouda Project

- Powerful
- Flexible

Main Founders:

- Michael Merrill – US government
- William Reus – US government

Main Contributors:

- https://github.com/Bears-R-Us/arkouda/graphs/contributors

NJIT New Jersey Institute of Technology

Arkouda: Dedication to **Michael H. Merrill**

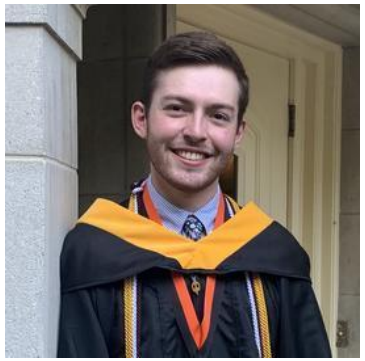
(June 2, 1964 ~ November 8, 2022)

“Mike was a dedicated civil servant. He was a Computer Scientist at the Department of Defense for 34 years and was recognized in 2022 with a Distinguished Civilian Service Medal. He loved computers and technology, especially high performance computing. Mike was a problem solver and innovative thinker; he was recognized for inspiring and leading numerous large projects over the course of his career. He loved to share his knowledge and mentored many colleagues over the years—sometimes calling them his kids, sometimes his minions, but always calling them his friend.”

NJIT New Jersey Institute of Technology

Arachne Developers

Zhihui Du
Research Scientist

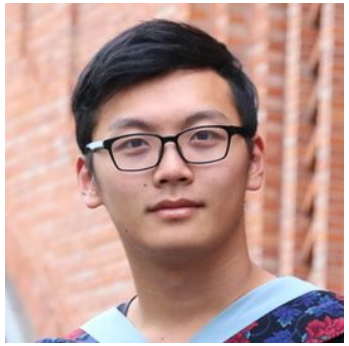


Oliver Alvarado Rodriguez
PhD Candidate

Naren Khatwani
MS student

Joseph Patchett
MS student (grad)

David Bader
Distinguished Professor



Fuhuan Li
PhD Student

Rajendra Prasad Patil
MS Student (grad)

Vanshika Agrawal
Undergrad Student

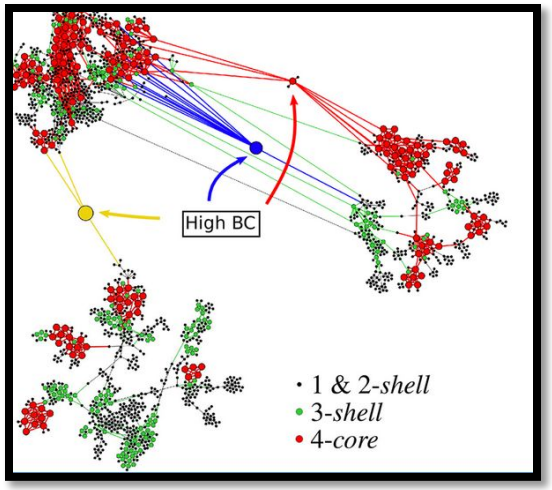
Davor Petrovikj
Undergrad Student

NJIT New Jersey Institute of Technology

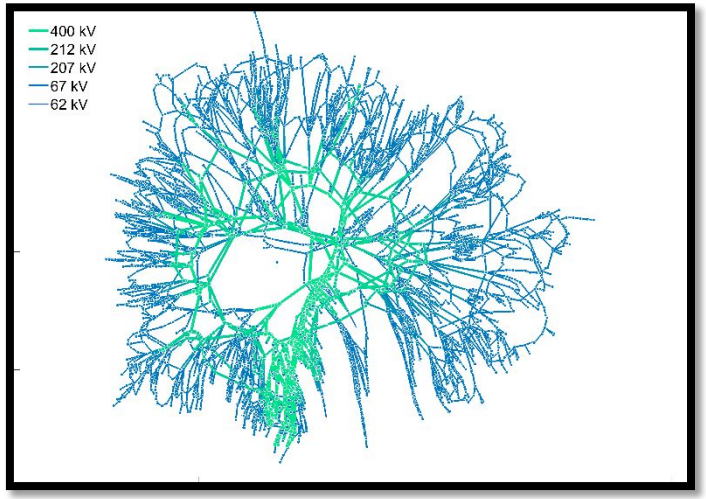
Motivation 1: Massive-Scale Real-World Graphs

- Public Health epidemic/pandemic contact tracing networks (COVID-19, monkey pox, flu(s))
- Sociology massive social media networks, email communication networks
- Bioinformatics protein and genome networks
- Neuroscience brain connectivity networks
- Urban Planning electrical power grid networks
- Scientometrics co-authorship networks
- Finance transaction networks

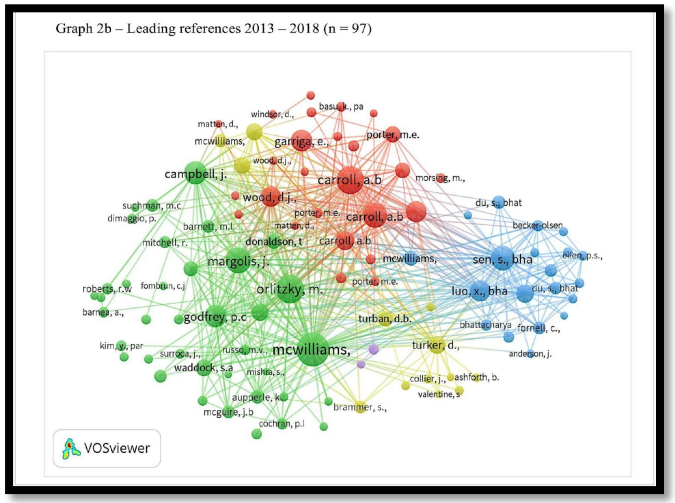
Motivation 2: Need for Graph Analytics



Analyzed infection network of COVID to find vertices with high betweenness centrality values that signified a super-spreader between communities [Serafino, Monteiro, et al. 2022].



Visualization of a high voltage transmission system to find the edges with maximum flow [Cuffe, 2016].

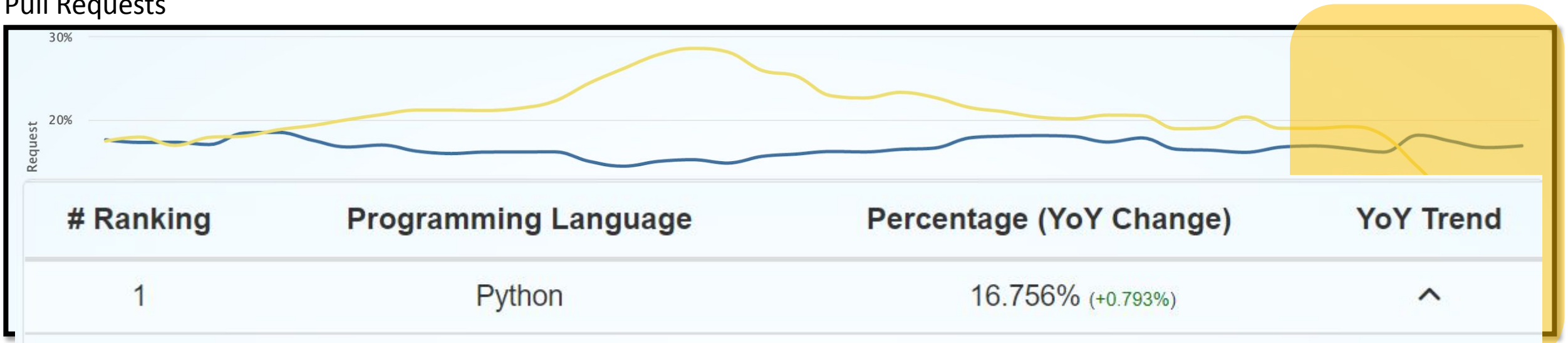


Finding the communities amongst authors in the area of corporate social responsibility [Ferramosca, Verona, 2019].

New Jersey Institute of Technology

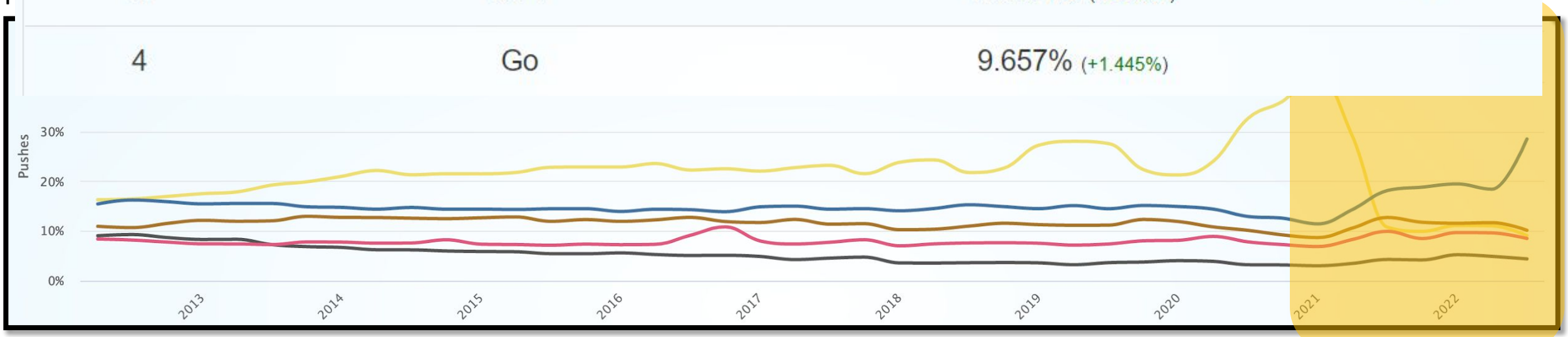
Motivation 3: Easy and Efficient Data Science

Pull Requests

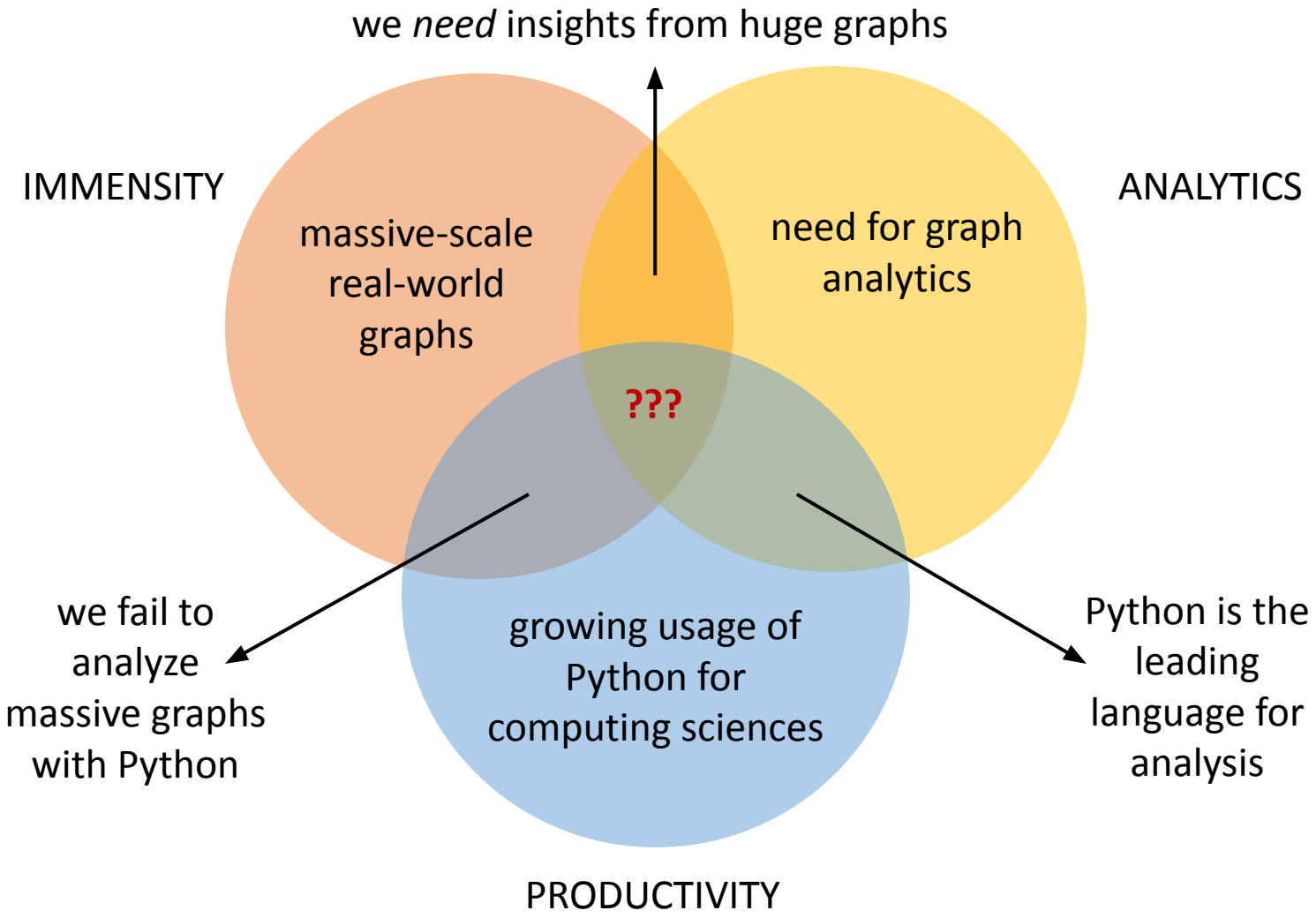


# Ranking	Programming Language	Percentage (YoY Change)	YoY Trend
1	Python	16.756% (+0.793%)	^
2	Java	11.005% (-1.988%)	^
3	C++	10.254% (+3.563%)	^
4	Go	9.657% (+1.445%)	

P

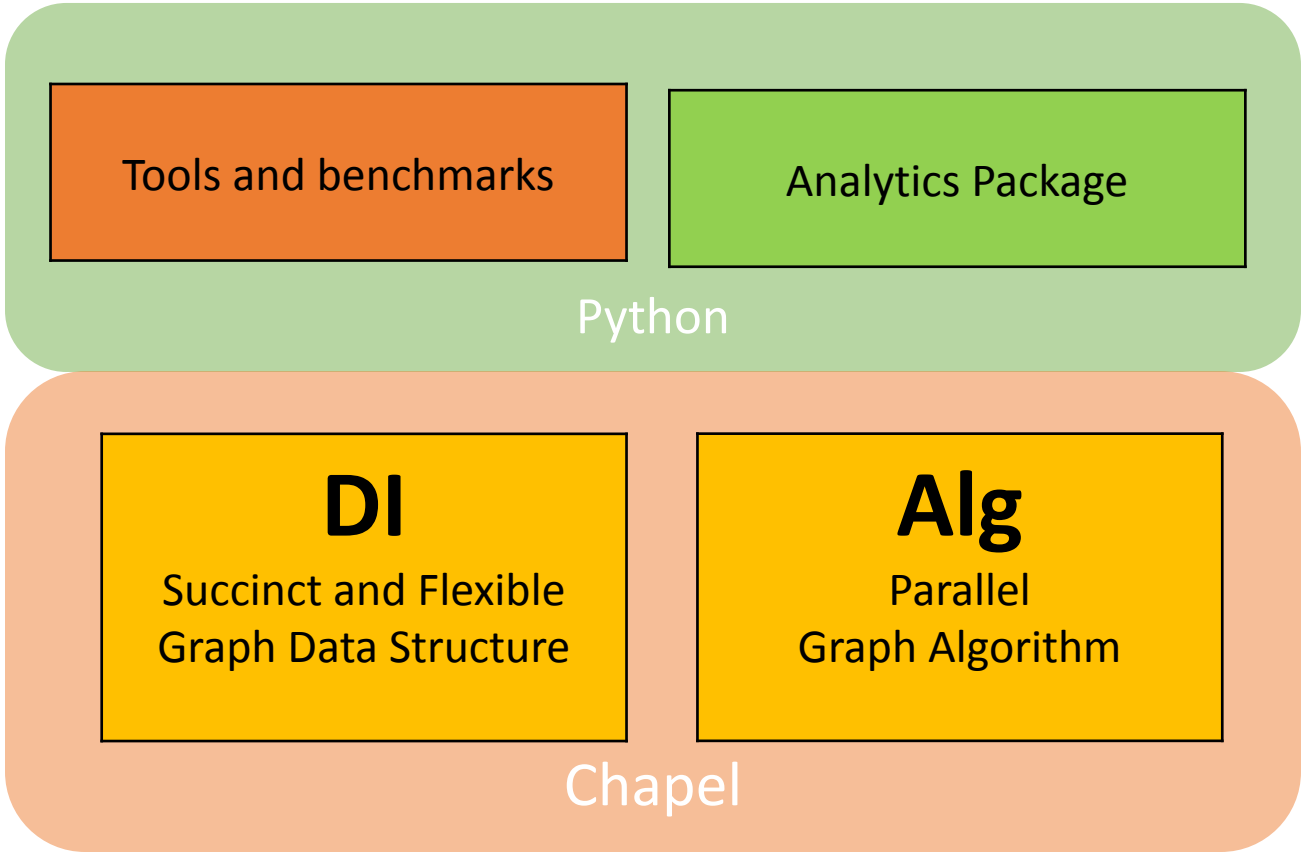


Motivations Overlap



NJIT New Jersey Institute of Technology

Arachne System

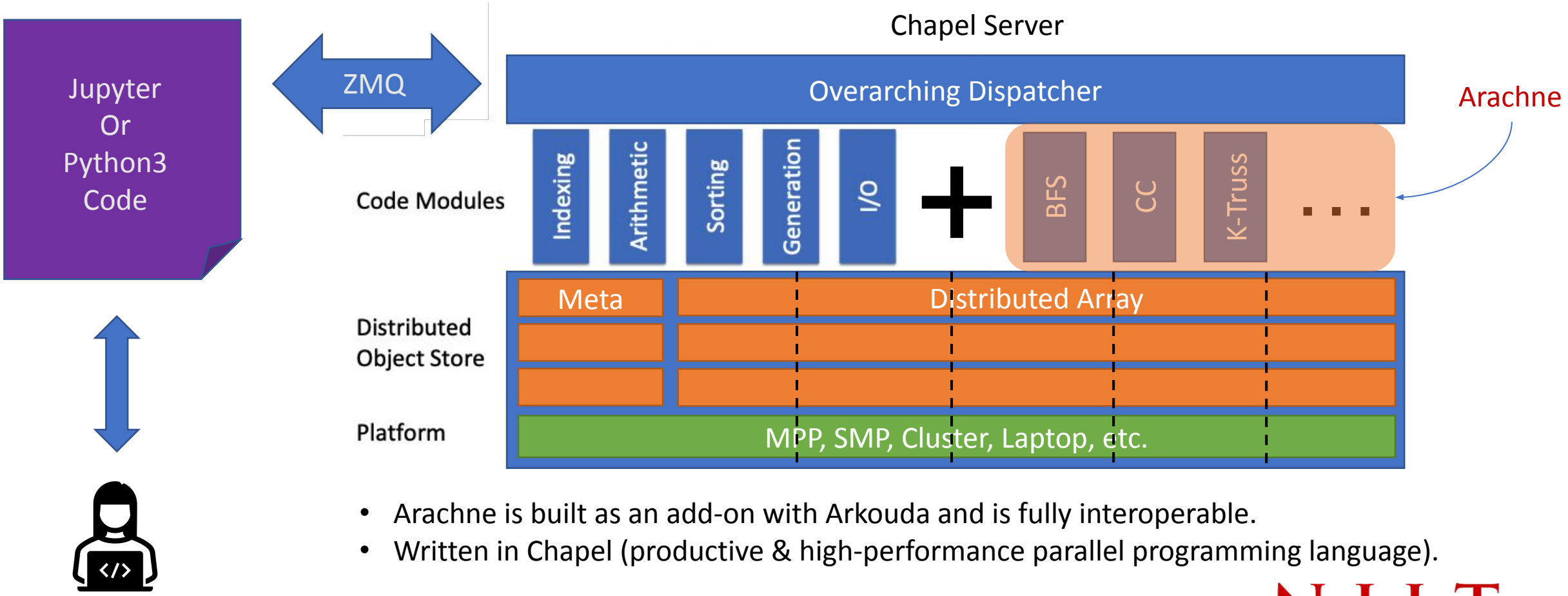


Interface

Core

New Jersey Institute of Technology

Integrate Arachne into Arkouda Framework



- Arachne is built as an add-on with Arkouda and is fully interoperable.
- Written in Chapel (productive & high-performance parallel programming language).

Image source: https://chapel-lang.org/CHIUW/2020/Reus.pdf

New Jersey Institute of Technology

How to use Arkouda and/or Arachne?

Use a Docker image.

Building from source. Works for both Arkouda and/or Arachne.

Using Arkouda from Docker

1. Install the latest version of Docker for Windows/MacOS/Linux.
2. Start Docker on your system.
3. Clone the Arkouda repository from (github.com/Bears-R-Us/arkouda) if you haven't already done so.
4. Install the environment for Arkouda into an Anaconda environment.
5. Run the command below to launch an Arkouda container with a mounted directory (github.com/Bears-R-Us/arkouda-contrib/tree/main/arkouda-docker).

```
# set env variables
export ARKOUDA_IMAGE_REPO=bearsrus
export ARKOUDA_VERSION=v2022.11.17

docker run -it --rm -p 5555:5555 --mount type=bind,source=/path/to/arkouda/resources/hdf5-testing,target=/app
$ARKOUDA_IMAGE_REPO/arkouda-smp-server:$ARKOUDA_VERSION
```

Building Arkouda and Arachne together from Source

1. Download and Build Chapel (chapel-lang.org/download.html).
2. Download Arkouda (github.com/Bears-R-Us/arkouda).
3. Download Arachne (github.com/alvaradoo/arkouda-contrib/tree/main/arachne).
   - That is the temporary location for Arachne, will be moving into arkouda-contrib maintained by the developers.
4. Follow the installation steps (github.com/alvaradoo/arkouda-contrib#installation)
5. Run the Arkouda server! (github.com/Bears-R-Us/arkouda#running-arkouda_server-toc)

How the Arkouda Server Looks Like

```
*********************************************************************************************
*********************************************************************************************
*                                                                                           *
*                          server listening on tcp://ee34776e7064:5555                      *
*                              arkouda server version = v2023.02.08                         *
*                                built with chapel version1.29.0                           *
*                                   memory limit = 3712075776                               *
*                                    bytes of memory used = 0                               *
*                                                                                           *
*********************************************************************************************
*********************************************************************************************
```

Connecting to Arkouda Server & Run Small Example

In [1]:
```
import arkouda as ak
```

```
    _         _                   _       
   / \   _ __| | _____  _   _  __| | __ _ 
  / _ \ | '__| |/ / _ \| | | |/ _` |/ _` |
 / ___ \| |  |   < (_) | |_| | (_| | (_| |
/_/   \_\_|  |_|\_\___/ \__,_|\__,_|\__,_|
```

```
Client Version: v2022.10.13+4.gb2b4b034
```

In [2]:
```
ak.connect()
```

```
/Users/alvaradoo/Research/arkouda/arkouda/client.py:199: RuntimeWarning: Version mismatch between client (v2022.10.
13+4.gb2b4b034) and server (v2022.11.17); this may cause some commands to fail or behave incorrectly! Updating arko
uda is strongly recommended.
  warnings.warn(
connected to arkouda server tcp://*:5555
```

In [3]:
```
f = ak.ones(10)
```

In [4]:
```
print(f)
```

```
[1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+
00 1.000000e+00]
```

Thank You ☺️

Questions?