Massive-scale Data Science
for IARPA Rapid Explanation, Analysis and Sourcing Online (REASON) program

David A. Bader

🐦 **@Prof_DavidBader**

http://www.cs.njit.edu/~bader

David A. Bader

Distinguished Professor and
Director, Institute for Data Science

- IEEE Fellow, ACM Fellow, SIAM Fellow, AAAS Fellow

- IEEE Sidney Fernbach Award

- 2022 inductee into University of Maryland's Innovation Hall of Fame, A. James Clark School of Engineering

- Recent Service:
  - White House's National Strategic Computing Initiative (NSCI) panel
  - Computing Research Association Board
  - Chair, NSF Committee of Visitors for Office of Advanced Cyberinfrastructure
  - NSF Advisory Committee on Cyberinfrastructure
  - Council on Competitiveness HPC Advisory Committee
  - IEEE Computer Society Board of Governors
  - IEEE IPDPS Steering Committee
  - Editor-in-Chief, ACM Transactions on Parallel Computing
  - Editor-in-Chief, IEEE Transactions on Parallel and Distributed Systems

- Over $186M of research awards

- 300+ publications, ≥ 13,000 citations, h-index ≥ 65

- National Science Foundation CAREER Award recipient

- Directed: Facebook AI Systems

- Directed: NVIDIA GPU Center of Excellence, NVIDIA AI Lab (NVAIL)

- Directed: Sony-Toshiba-IBM Center for the Cell/B.E. Processor

- Founder: Graph500 List benchmarking "Big Data" platforms

- Recognized as a "RockStar" of High Performance Computing by InsideHPC in 2012 and as HPCwire's People to Watch in 2012 and 2014.

Prior IARPA and DARPA performer: e.g., HPCS, UHPC, ADAMS, PERFECT, HIVE, SDH

NJIT
New Jersey Institute of Technology

Massive-Scale Data Science:
The Arkouda Framework with Arachne (Graph Analytics)
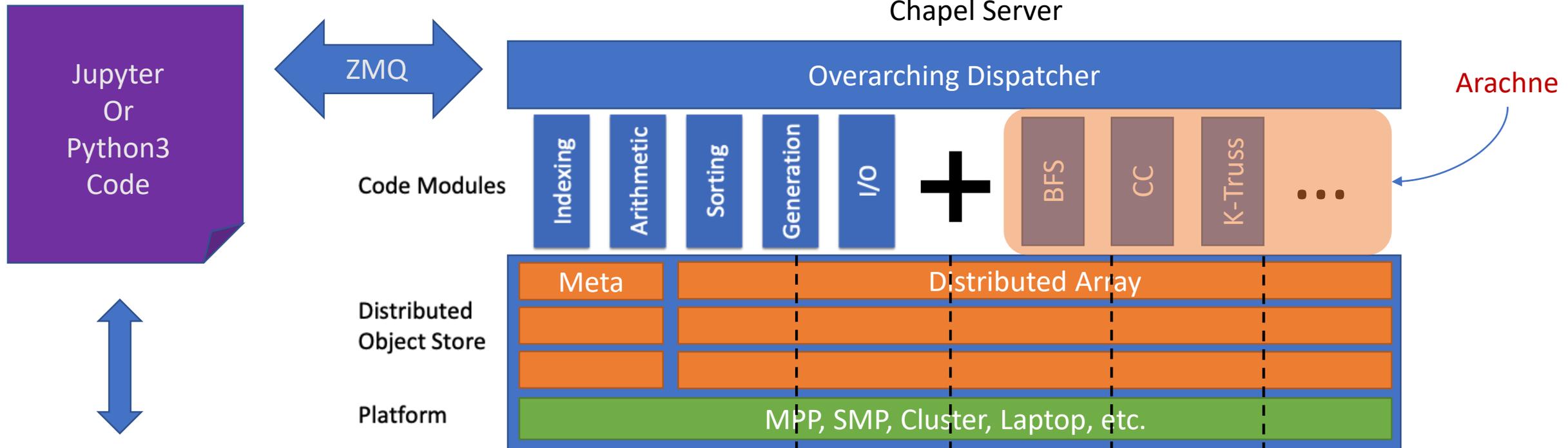
- Arachne is built as an add-on with Arkouda and is fully interoperable.
- Where can I get it?
  - Software: https://github.com/mhmerrill/arkouda
  - Our Contribution: https://github.com/Bader-Research/arkouda/tree/streaming

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

Major Contributions

- Arachne, a large-scale graph analysis framework, extends Arkouda for productive graph analysis. Arachne is built on a novel sparse graph data structure and includes BFS, connected components, truss analytics, Jaccard coefficients, triangle counting, centrality, and more (property graphs and community detection planned for future).

- Arachne leverages productivity through Python with high performance backed by Chapel.

- Arachne, Arkouda, Chapel are all open-source.
  - https://github.com/Bears-R-Us/arkouda-njit
  - https://github.com/Bears-R-Us/arkouda
  - https://github.com/chapel-lang/chapel

- Experimental results on real-world and synthetic graphs demonstrate that Arachne works for graphs with billions to trillions of edges.

Publications

- **Oliver Alvarado Rodriguez, Zhihui Du, Joseph Patchett, Fuhuan Li, David Bader (2022). Arachne: An Arkouda Package for Large-Scale Graph Analytics. IEEE HPEC.**

- 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