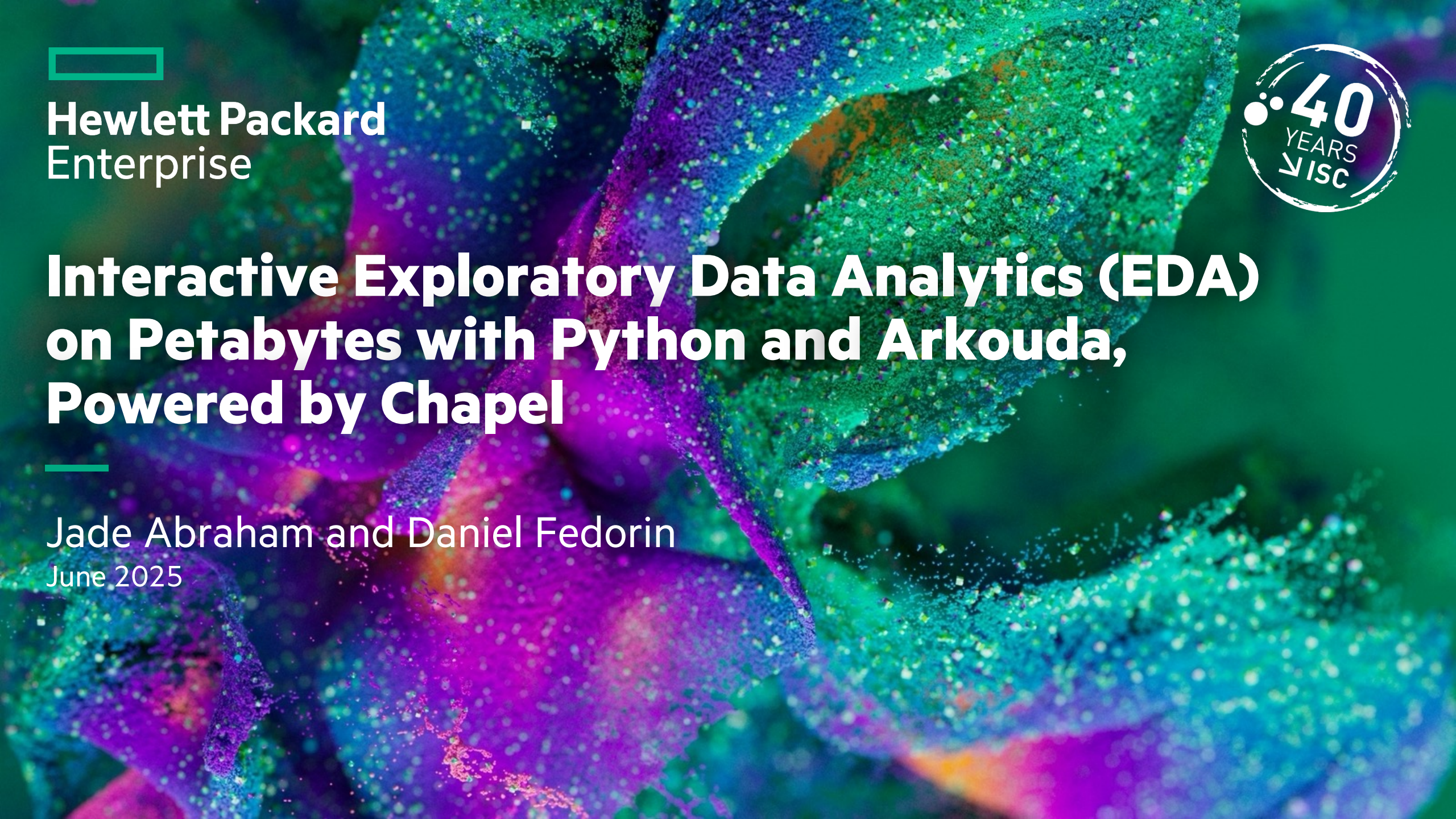
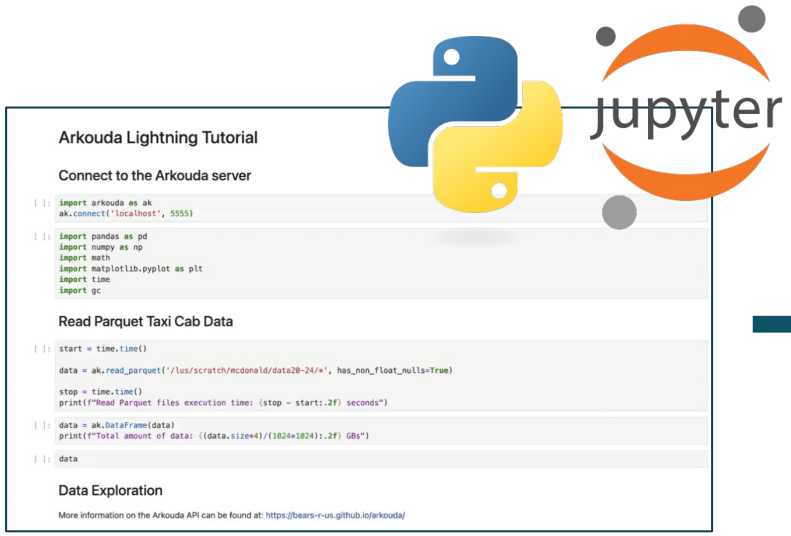


Interactive Exploratory Data Analytics (EDA) on Petabytes with Python and Arkouda, Powered by Chapel

Jade Abraham and Daniel Fedorin

June 2025

Arkouda Puts Massive-Scala Data Analysis at your Fingertips

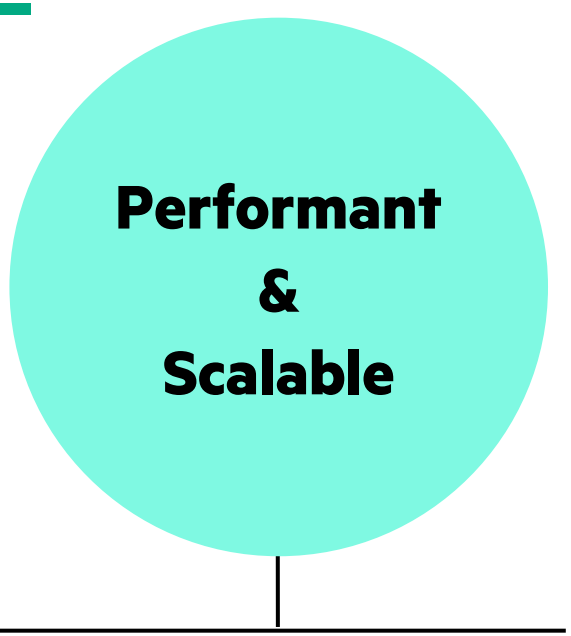


Local Dev. Environment

Arkouda

HPC Systems

Arkouda Enables HPC from Python



Interactive Rates

Operations run in seconds

Massive Scales

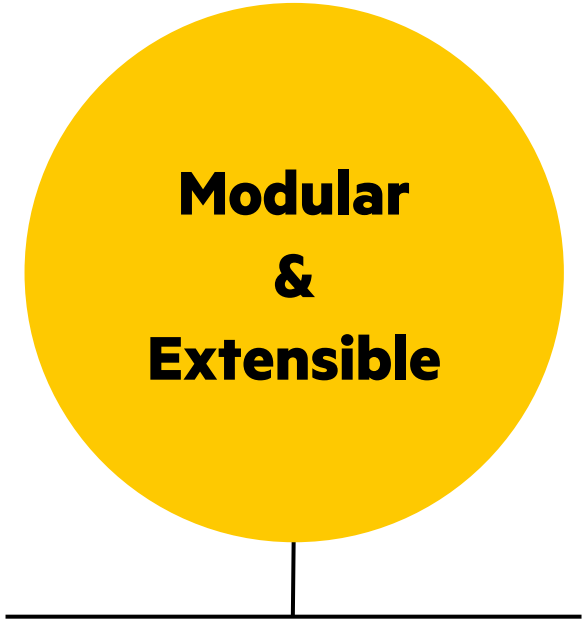
100s of TBs using 1000s of nodes

Python Interface for Client

Familiar, interactive, Jupyter-ready

Chapel-powered Server

Runs on supercomputer, cluster, cloud



Modules for Different Uses

Graphs, visualizations, and more

Open-Source

Developed under MIT License

No other tool provides Exploratory Data Analysis (EDA) at these scales

Arkouda Demo

Client

- Jupyter on laptop

Server

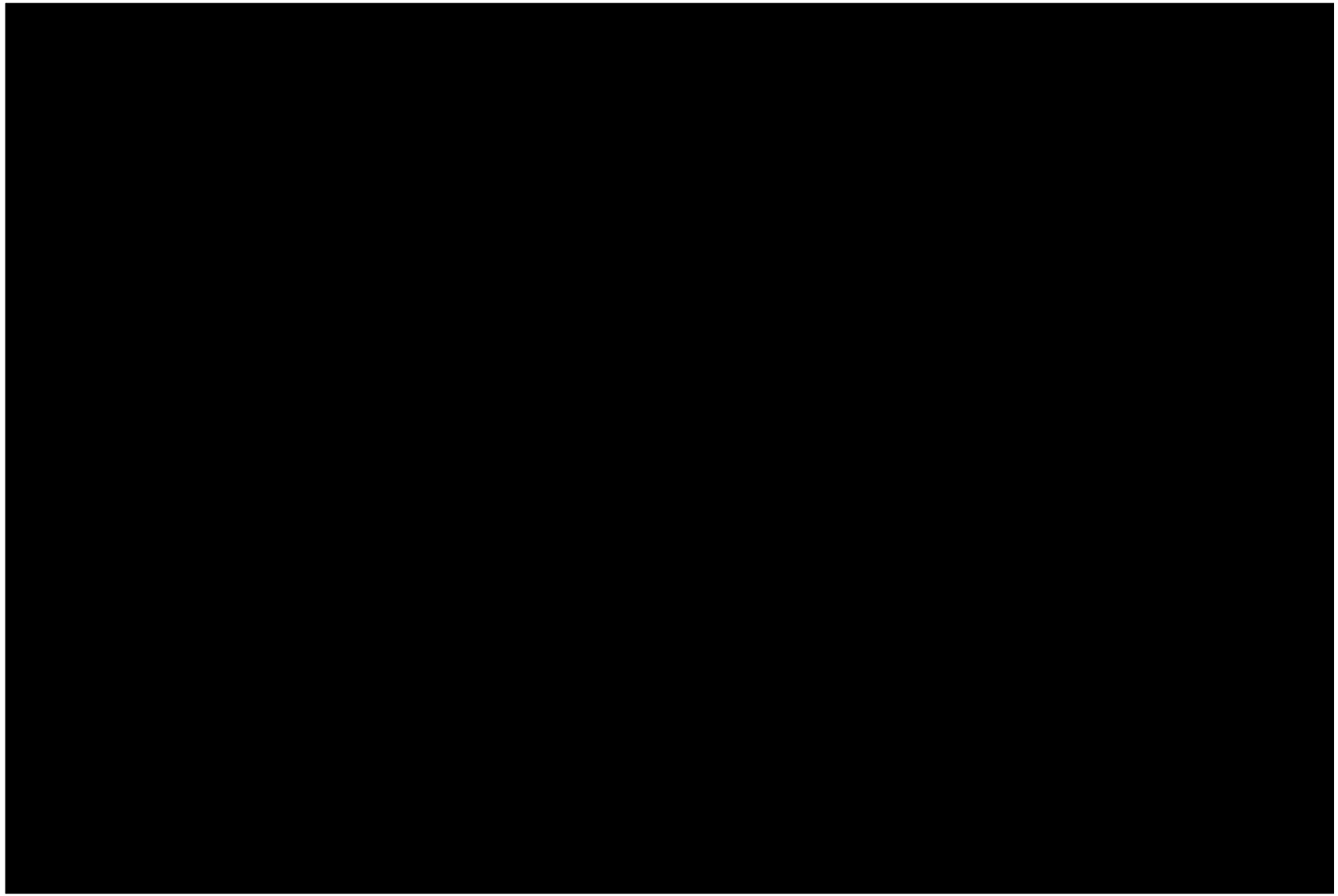
- Running on 8 nodes of Cray XC

Dataset

- Too large for a laptop

Operations

- Histogram
- Group-by
- Visualization w/ matplotlib



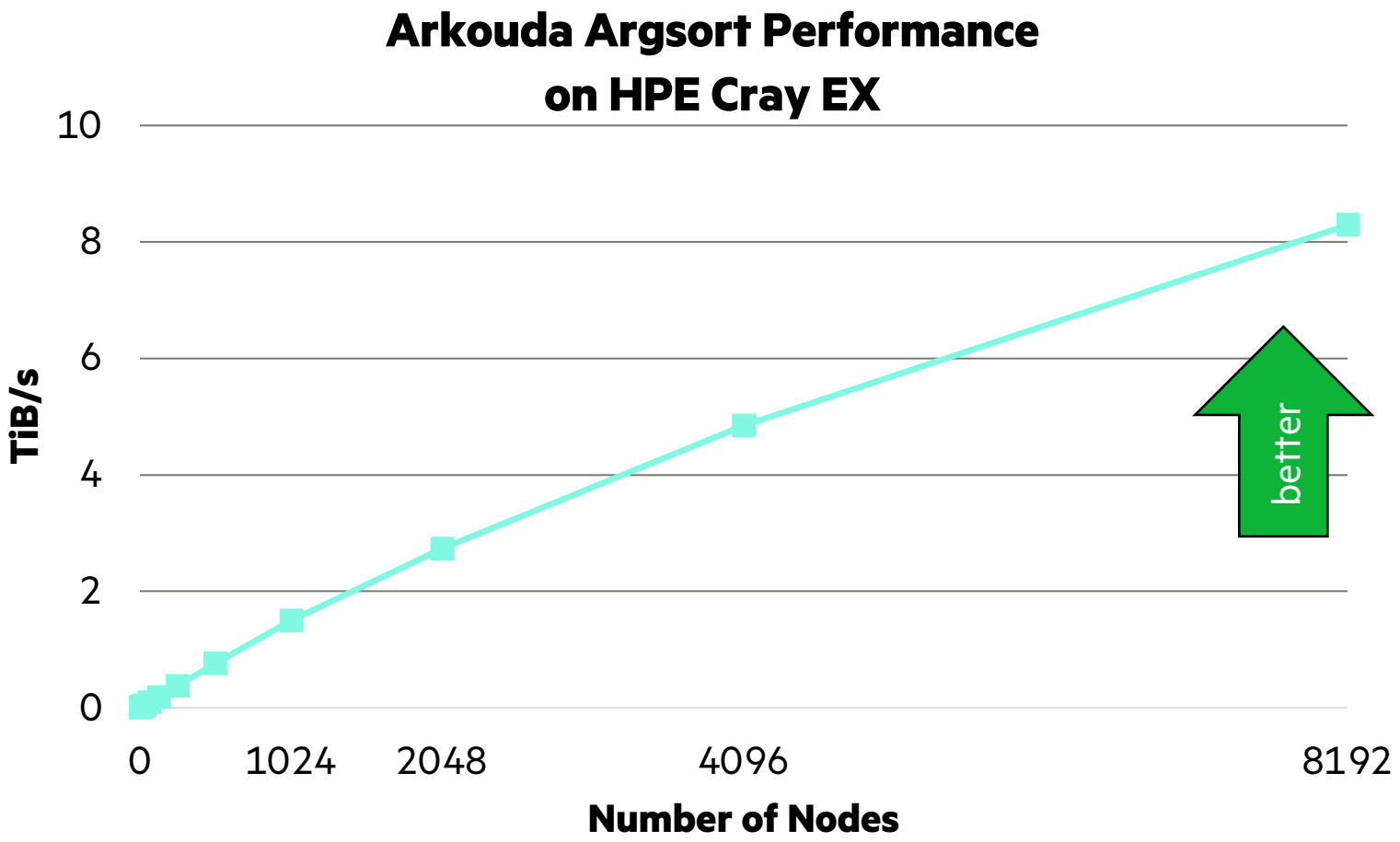
Radix Sort in Arkouda Scaling to 8 TiB/s on 8K Nodes

Slingshot-11 network (200 Gb/s)

8192 compute nodes

256 TiB of 8-byte values

~8500 GiB/s (~31 seconds)



A notable performance achievement in ~100 lines of Chapel

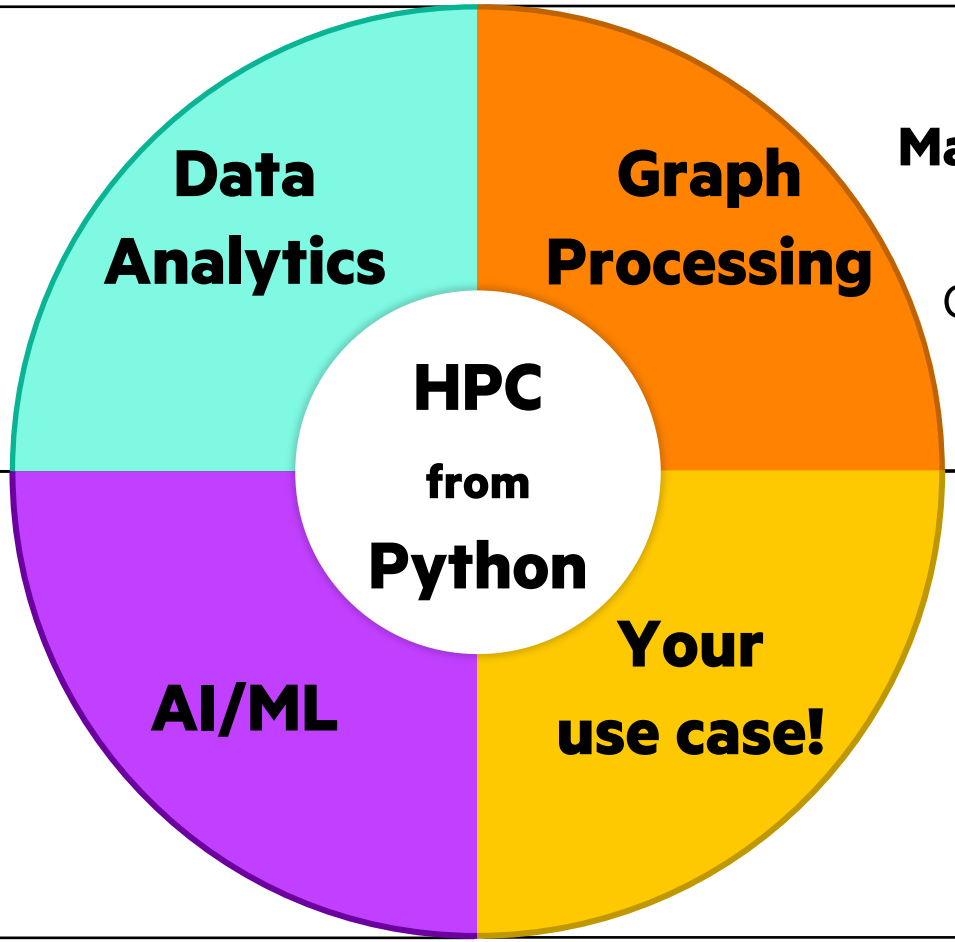
What can Arkouda do?

Scalable Operations

sorting, groupby, stats…

Performant I/O

Parquet, HDF5…



Massive-scale real-world graphs

Arachne: Developed by NJIT

Graph infrastructure for data science

Distributed Inference

Chapel AI Library (ChAI)

Open-source, under development

Coming Soon!

Supercomputer telemetry

Earth sciences with Pangeo

Powered by the Chapel Parallel Programming Language

Chapel is a language designed for productive parallel programming, particularly on large-scale systems. Chapel is …

Easy to Use

"*We ask students at the master's degree to do stuff that would take 2 years and they do it in 3 months.*" Eric Laurendeau, Professor of Mechanical Engineering

Portable

HPE Cray EX, HPE Apollo, Cray XC, *nix systems, Mac, NVIDIA and AMD GPUs

Fast & Scalable

Achieved 8,500 GiB/s when sorting 256 TiB in 31 seconds on 8192 HPE Cray EX Nodes

GPU-Ready

Real-world applications were ported on GPUs with few changes, and run on leadership-class systems such as Frontier and Perlmutter

Open source

Team at HPE actively interacts with Chapel community at chapel-lang.org

Thank you