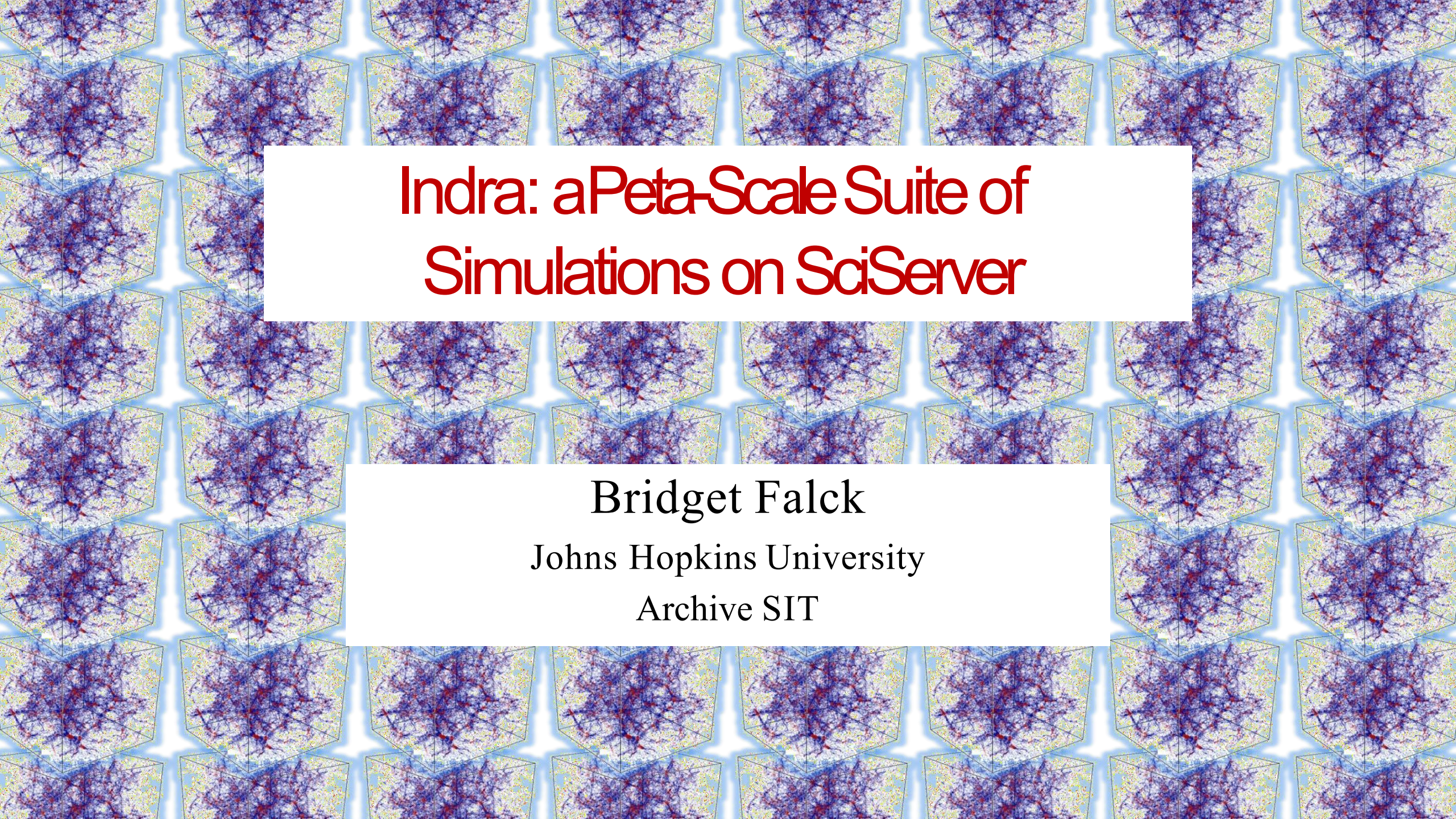


# Indra: a Peta-Scale Suite of Simulations on SciServer

Bridget Falck

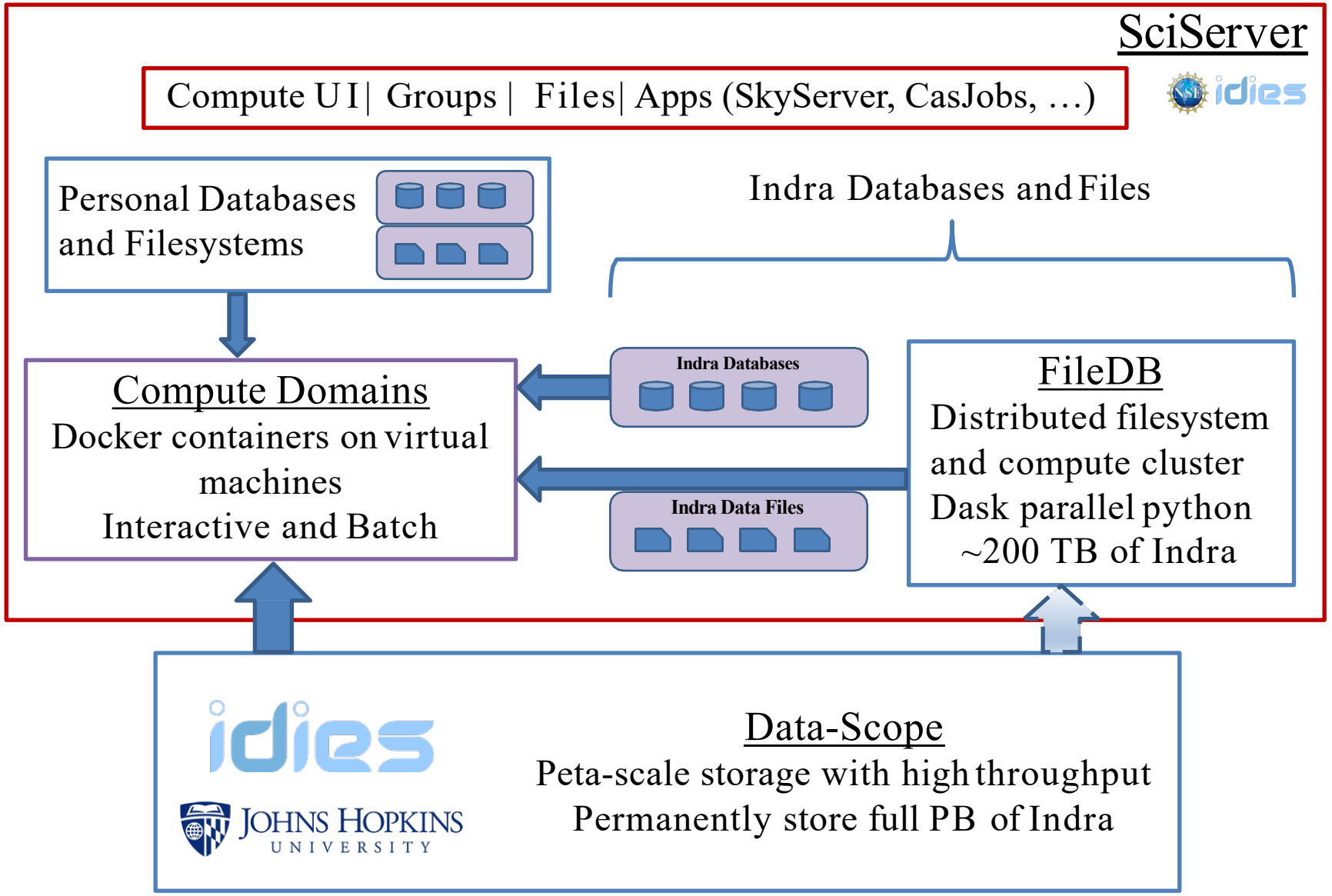
Johns Hopkins University

Archive SIT

# The Indra Simulations

- Suite of 384 simulations with the same cosmology (BF+2101.03631)
  - Each a 1 Gpc/h-sided box with $1024^3$ dark matter particles, run with L-Gadget
  - Output: 64 snapshots of particle positions and velocities, FOF/SUBFIND halo catalogs, and 505 time-steps of Fourier-space density grids
- 750 TB of data, available to the public and *computationally-accessible* via SciServer
  - Ensemble averages and covariances, conditional and extreme statistics, mock galaxy catalogs and lightcones, etc.
  - Test-bed for new data architectures and analysis tools

# Indra Infrastructure



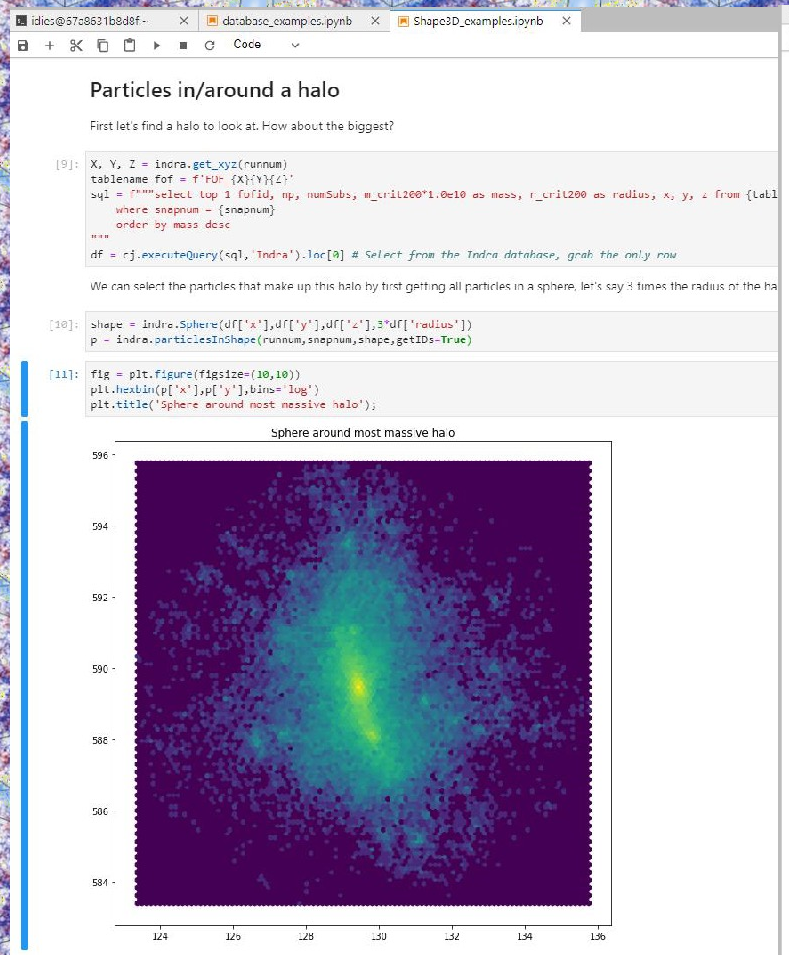

- Permanent read-only storage connected to compute domain
- Distributed filesystem for parallel computation
- Leverage relational databases
- Accessible through SciServer with its collaboration tools and hosted astrophysical datasets

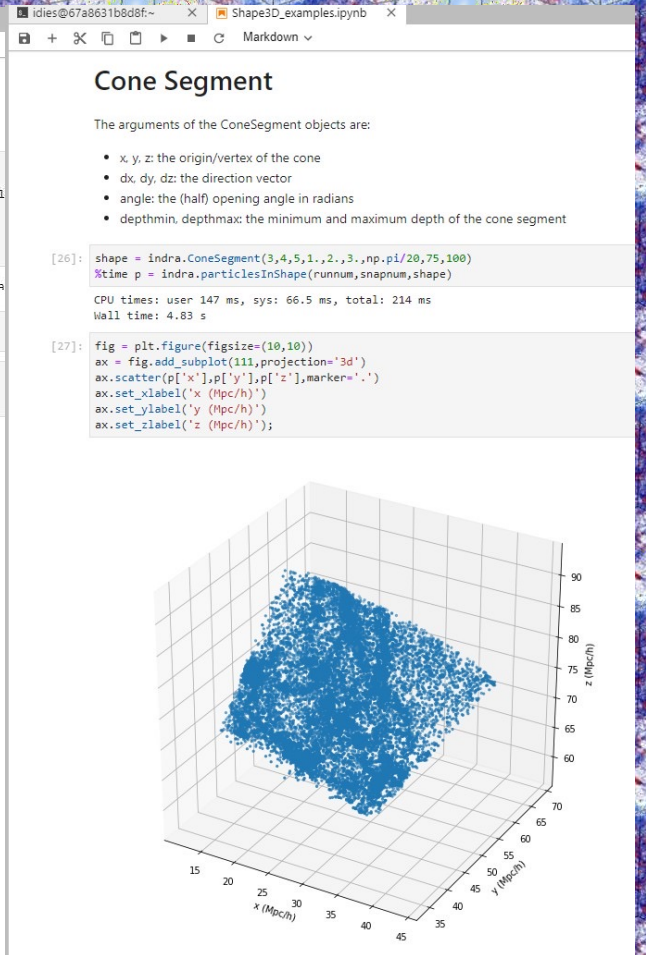
# Tools for efficient data access

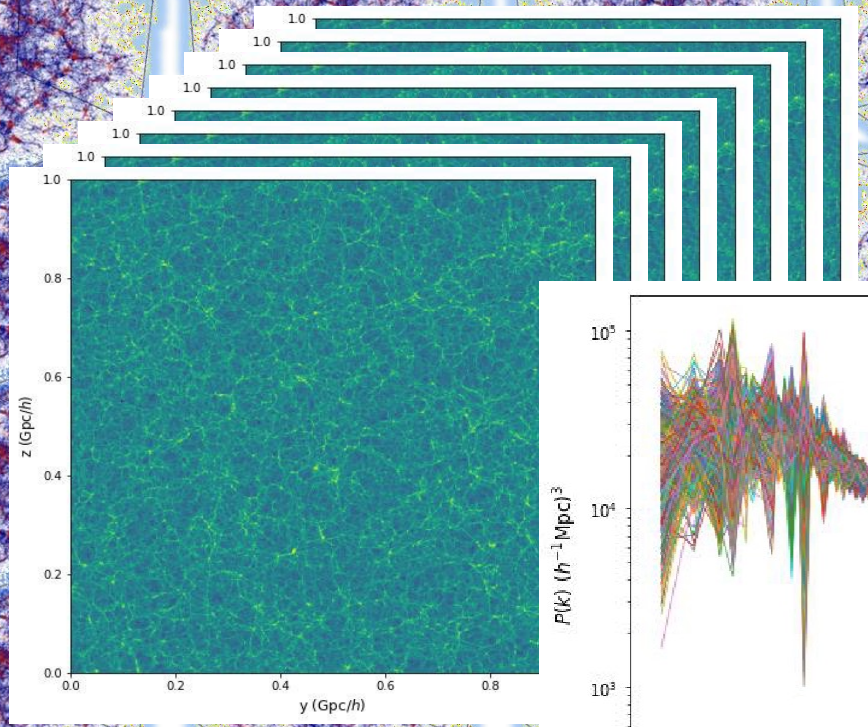
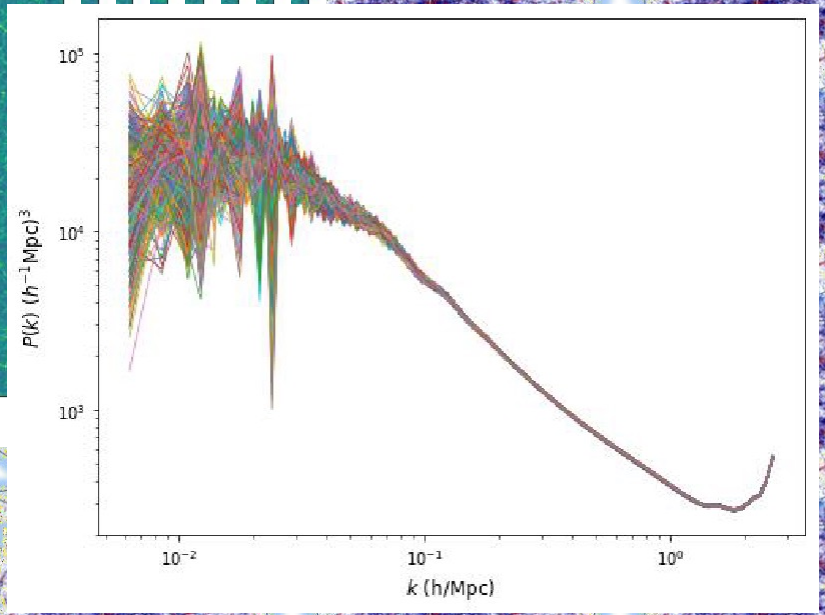
- Python library to read and interact with data
  - Don't assume expert users
  - Make it easier for experts
  - Hide the file system

  github.com/bfalck/indra-tools

- Ex: Shape3D
  - Leverage spatial indexing in relational database
  - Select data within shapes
  - Compute halo density profiles or light cones



# Infrastructure that enables heavy computation

- FileDB distributed filesystem with a Dask parallel python cluster
- Submit compute tasks and collect output from within Jupyter notebook
- 448 Cloud-In-Cell density grids calculated in 2 hours!
  - 481 billion particles total