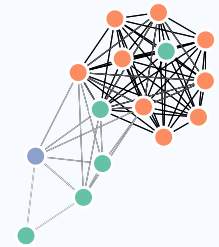# ANDERSON BANIHIRWE

I contribute to and maintain several libraries within the open source scientific Python stack, particularly around improving scalability of Python tools in order to handle terabyte-scale datasets on HPC and cloud platforms.

🎓 EDUCATION

2018 | 2014

B.S., Computer Systems Engineering
University of Arkansas at Little Rock

📍 Little Rock, AR

💼 PROFESSIONAL EXPERIENCE

present | 2020-10

Software Engineer ‖
National Center for Atmospheric Research

📍 Boulder, CO

· Created jupyter-forward, a Jupyter Lab port forwarding utility that simplifies running jupyter on remote resources.
· Served as a core developer of xarray, an open source library for working with multidimensional labeled datasets and arrays in Python.

2020-9 | 2018-10

Software Engineer |
National Center for Atmospheric Research

📍 Boulder, CO

· Lead the intake-ESM project, a Python data cataloguing package for exploring and ingesting earth system model data sets.
· Contributed to the core software stack powering the Pangeo Project. Some of the projects I contributed to include: xarray, dask.
· Assisted with the development and deployment of live (virtual or in-person) and online/self-paced education material.

2018-09 | 2018-05

Software Developer Intern
Quansight

📍 Austin, TX

· Developed xndframes[1], a Pandas ExtensionDtype/Array backed by xnd[2], a container type that maps most Python values relevant for scientific computing directly to typed memory.
· Worked on integrating cuDF[3] - GPU dataframe library with Apache Arrow[4] library.

2018-04 | 2017-11

Data Science Intern
First Orion

📍 Little Rock, AR

· Built scoring, predictive models with Scikit-learn, Dask, and Apache Spark using First Orion's proprietary telecommunication data.

2017-08 | 2017-05

Research Intern
National Center for Atmospheric Research

📍 Boulder, CO

· Developed spark-xarray[5], a Python package that integrates PySpark and xarray for climate data analysis.

View this CV online with links at *cv.andersonbanihirwe.dev*

CONTACT

✉ axbanihirwe@gmail.com

 github.com/andersy005

🔗 blog.andersonbanihirwe.dev

in linkedin.com/in/andersy005

Last updated on 2020-12-16.

SELECTED PUBLICATIONS, POSTERS, AND TALKS

2020-11 — **Cloud-Native Repositories for Big Scientific Data[6]**

Computing in Science and Engineering

- Authored with Ryan Abernathey, Tom Augspurger, et al.

2020-10 — **Pangeo Benchmarking Analysis: Object Storage vs. POSIX File System[7]**

Fifth International Parallel Data Systems Workshop @ SC 20

- Authored with Haiying Xu, Kevin Paul

2020-01 — **The Pangeo Ecosystem: Interactive Computing Tools for the Geosciences: Benchmarking on HPC[8]**

2019 Supercomputing Conference Workshop on Interactive High-Performance Computing

- Authored with Tina Erica Odaka, Guillaume Eynard-Bontemps, Aurelien Ponte, Guillaume Maze, Kevin Paul, Jared Baker, Ryan Abernathey.

2020-09 — **Zarr: chunked, compressed, multidimensional arrays[9]**

2020 Cloud Native Geospatial Outreach Day   📍 Online

- Invited talk about Zarr[10], an open source data format for the storage of chunked, compressed, multidimensional arrays.

2020-07 — **Intake-ESM – Making It Easier To Consume Climate and Weather Data[11]**

2020 ESIP Summer Meeting   📍 Online

- Invited talk about intake-esm, an intake plugin for working with Earth System Model (ESM) datasets.

2019-01 — **Perceptual Judgments to Detect Computer Generated Forged Faces in Social Media[12]**

IAPR Workshop on Multimodal Pattern Recognition of Social Signals in Human-Computer Interaction

- Authored with Suzan Anwar, Mariofanna Milanova, Mardin Anwer.

2019-07 — **Interactive Supercomputing with Dask and Jupyter[13]**

2019 Scientific Computing with Python conference   📍 Austin, TX

- Contributed talk about Dask and Jupyter.

2018-04 — **Beyond Matplotlib – Tutorial: Building Interactive Climate Data Visualizations with Bokeh and Friends[14]**

2018 UCAR Software Engineering Assembly conference   📍 Boulder, CO

- Contributed tutorial about interactive visualization with Python.

2018-01 — **PySpark for "Big" Atmospheric Data Analysis**

Eighth Symposium on Advances in Modeling and Analysis Using Python

📍 Austin, TX

- Contributed talk about spark-xarray[15].