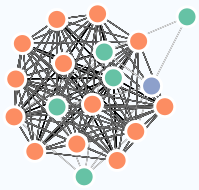
ANDERSON BANIHIRWE

I contribute to and maintain several libraries within the open source scientific Python stack, particularly around improving scalability of Python tools in order to (1) handle large scale datasets on High Performance Computing and Cloud Computing platforms and (2) move the open science paradigm forward.



EDUCATION

2018 | 2014

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

PROFESSIONAL EXPERIENCE

current | 2022-4

Software/Data Engineer
CarbonPlan — San Francisco, CA

- Spearheaded cross-functional coordination within the CarbonPlan team to support internal research and policy initiatives, enhancing and iterating on data tools. Led development on key projects, including OffsetsDB and the Leap Data Catalog.
- Contributed to the Pangeo-Forge initiative by developing software tools and data processing pipelines for producing a wide range of analysis-ready cloud-optimized (ARCO) climate data
- Contributed to the development of existing open source scientific Python projects that are used at CarbonPlan and by the Pangeo community. Examples include the Xarray, Dask, Intake projects.

2022-03 | 2018-10

Software Engineer
National Center for Atmospheric Research — Boulder, CO

- Maintained and enhanced the core software stack driving the Pangeo project, with key contributions to projects such as Dask and Intake.
- Developed and maintained Xarray, an open source library for working with multidimensional, labeled datasets and arrays in Python.
- Created and maintained intake-ESM, a Python data cataloguing package for exploring and ingesting earth system model data sets.
- Developed and delivered weekly, live (virtual and in-person), self-paced technical tutorials to NCAR scientists and their collaborators.

2018-09 | 2018-05

Software Developer Intern
Quansight — Austin, TX

- Created and 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 to enhance data processing capabilities.

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

CONTACT

axbanihirwe@gmail.com
github.com/andersy005
blog.andersonbanihirwe.dev
linkedin.com/in/andersy005

Last updated on 2024-06-11.

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.

SELECTED PUBLICATIONS, POSTERS, AND TALKS

2022-01

Building Tools for the Scientific Python Community

12th Symposium on Advances in Modeling and Analysis Using Python at 2022 AMS Annual Meeting — Online

- Invited Keynote talk.

2021-05

The current State of Deploying Dask on HPC Systems

2021 Dask Developer Summit — Online

- Contributed talk.

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]

5th International Parallel Data Systems Workshop at 2020 Supercomputing Conference

- 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.

2020-10

Pangeo Use Case: Analyzing Initialized Climate Prediction System Datasets with climpred[9]

NOAA's 45th Climate Diagnostics & Prediction Workshop — Online

- Invited talk about climpred[10], a Python package for weather and climate forecasts.

2020-09 **Zarr: chunked, compressed, multidimensional arrays**[11]
2020 Cloud Native Geospatial Outreach Day — Online
- Invited talk about Zarr[12], 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**[13]
2020 ESIP Summer Meeting — Online
- Invited talk about intake-esm, an intake plugin for working with Earth System Model (ESM) datasets.

2020-02 **Dask and Pangeo**
2020 Dask Developer Summit — Washington, D.C.
- Invited talk.

2019-01 **Perceptual Judgments to Detect Computer Generated Forged Faces in Social Media**[14]
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**[15]
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**[16]
2018 UCAR Software Engineering Assembly conference — Boulder, CO
- Contributed tutorial about interactive visualization with Python.

2018-01 **PySpark for "Big" Atmospheric Data Analysis**
8th Symposium on Advances in Modeling and Analysis Using Python at 2018 AMS Annual Meeting — Austin, TX
- Contributed talk about spark-xarray[17].

LINKS

1: https://github.com/xnd-project/xndframes
2: https://github.com/xnd-project
3: https://github.com/rapidsai/cudf
4: https://arrow.apache.org/
5: https://ncar.github.io/PySpark4Climate/
6: https://www.authorea.com/doi/full/10.22541/au.160443768.88917719
7: https://doi.org/10.31223/X5ZW2T
8: https://doi.org/10.1007/978-3-030-44728-1_12

9: https://talks.andersonbanihirwe.dev/climpred-cdpw-2020.html
10: https://github.com/pangeo-data/climpred
11: https://talks.andersonbanihirwe.dev/zarr-cloud-native-geospatial-2020.html
12: https://github.com/zarr-developers
13: https://talks.andersonbanihirwe.dev/intake-esm-esip-2020.html
14: https://www.researchgate.net/profile/Mariofanna_Milanova/publication/333414231_Perceptual_Judgments_to_Detect_Computer_Generated_Forged_Faces_in_Social_Media/links/5e2c963092851c3aaddac2f5/Perceptual-Judgments-to-Detect-Computer-Generated-Forged-Faces-in-Social-Media.pdf
15: https://youtu.be/vhawO8fgD64
16: https://sea.ucar.edu/event/beyond-matplotlib-building-interactive-climate-data-visualizations-bokeh-and-friends
17: https://ncar.github.io/PySpark4Climate/sparkxarray/overview/