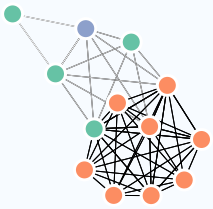
ANDERSON BANIHIRWE

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

current | 2018-10

Software Engineer

National Center for Atmospheric Research — Boulder, CO

- Assisted with the development, implementation, and maintenance of software contributions to Open Source packages –namely dask, xarray, intake, jupyter — used by the Pangeo community as they pertain to NCAR systems or data.
- Assisted with the development and maintenance of Pangeo-related installations and deployments on premise High Performance Computing (HPC) systems or commercial cloud that access NCAR resources.
- Assisted with the development and deployment of live (virtual or in-person) and online/self-paced education material, including training websites, Jupyter notebook examples, and teaching modules.

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.
- Worked closely with a customer to port existing Postgres code base to a Dask based workflow.

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

linkedin.com/in/andersy005

Last updated on 2020-09-18.

SELECTED PUBLICATIONS, POSTERS, AND TALKS

2020

The Pangeo Ecosystem: Interactive Computing Tools for the Geosciences: Benchmarking on HPC[6]

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

2020 Cloud Native Geospatial Outreach Day — Online

- Invited talk about Zarr[7], an open source data format for the storage of chunked, compressed, multidimensional arrays.
- Recorded talk: https://www.youtube.com/watch?v=cOMkgQssVPk&list=PL3QzFgBMGnbQWbW-V09AzSfCbnf6Q87Rq&index=4&t=1471s
- Slides: https://talks.andersonbanihirwe.dev/zarr-cloud-native-geospatial-2020.html

2020-07

Intake-ESM – Making It Easier To Consume Climate and Weather Data

2020 ESIP Summer Meeting — Online

- Invited talk about intake-esm, an intake plugin for working with Earth System Model (ESM) datasets.
- Recorded talk: https://youtu.be/OpzF6IwIHRA?t=1400
- Slides: https://talks.andersonbanihirwe.dev/intake-esm-esip-2020.html

2020-06

Intake / Pangeo Catalog: Making It Easier To Consume Earth's Climate and Weather Data[8]

2020 EarthCube Annual Meeting — Online

- Contributed Jupyter notebook about Pangeo[9]'s data cataloging efforts.

2019

Perceptual Judgments to Detect Computer Generated Forged Faces in Social Media[10]

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[11]

2019 Scientific Computing with Python conference — Austin, TX

- Contributed talk about Dask and Jupyter.
- Recorded talk: https://youtu.be/vhawO8fgD64
- Slides: https://andersonbanihirwe.dev/talks/dask-jupyter-scipy-2019.html

2018-04

Beyond Matplotlib - Tutorial: Building Interactive Climate Data Visualizations with Bokeh and Friends[12]

2018 UCAR Software Engineering Assembly conference — Boulder, CO

- Contributed tutorial about interactive visualization with Python.
- Tutorial materials: https://github.com/andersy005/beyond-matplotlib-tutorial-sea-2018

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[13].
- Recorded Talk: https://ams.confex.com/ams/98Annual/webprogram/Paper334546.html
- Slides: https://opensky.ucar.edu/islandora/object/conference%3A3443

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://doi.org/10.1007/978-3-030-44728-1_12
7: https://github.com/zarr-developers
8: https://github.com/earthcube2020/ec20_banihirwe_etal
9: https://pangeo.io/
10: https://www.researchgate.net/profile/Mariofanna_Milanova/publication/333414231_Perceptual_Judgments_to_Detect_Computer_Ge
Judgments-to-Detect-Computer-Generated-Forged-Faces-in-Social-Media.pdf
11: https://youtu.be/vhawO8fgD64
12: https://sea.ucar.edu/event/beyond-matplotlib-building-interactive-climate-data-visualizations-bokeh-and-friends
13: https://ncar.github.io/PySpark4Climate/sparkxarray/overview/