Anderson Banihirwe

https://andersy005.github.io/

Email : axbanihirwe@gmail.com
Github : andersy005
LinkedIn: andersy005

Education

- **University of Arkansas at Little Rock** — Little Rock, AR
  Bachelor of Science in (Computer) Systems Engineering; GPA: 3.83/4.00 — *Aug 2014 – May 2018*

Professional Experience

- **Quansight**
  Intern — *June 2018 – Present*
  - **xnd**: Worked on xnd, a library for refactoring of NumPy capabilities to low-level libraries and high-level interfaces.
  - **GPU dataframes support**: Worked on integrating pygdf (a Python interface to access and manipulate the GPU DataFrame) with Apache Arrow library.
- **First Orion** — Little Rock, AR
  Data Scientist Intern — *Dec 2017 – May 2018*
  - **Machine Learning**: Designed and built scoring, predictive models with Scikit-learn using First Orion' s proprietary telecommunication data.
  - **Data Processing**: Identified patterns and characteristics within First Orion's data warehouses using Dask, Apache Spark, Pandas.
- **National Center for Atmospheric Research** — Boulder, CO
  Research Intern — *May 2017 – Aug 2017*
  - **Installation**: Installed Apache Spark v2.2 on both Cheyenne and Yellowstone Supercomputers.
  - **Schedulers**: Cleaned/fixed Spark launch bash scripts that work with the LSF/PBS schedulers.
  - **spark-xarray**: Wrote spark-xarray, a python package that integrates PySpark and xarray for Climate Data Analysis.
  - **Jupyter notebooks contribution**: Contributed Jupyter notebooks and scripts using Apache Spark to NCAR's Coupled Model Intercomparison Project (CMIP) Analysis Platform.
  - **Documentation**: Documented research work at https://ncar.github.io/PySpark4Climate/

Selected Projects

- **Spark-xarray**: Open source python library built on top of PySpark - Spark Python API and xarray for climate data analysis. https://github.com/andersy005/spark-xarray
- **Dask**: Contributed examples to dask-examples repository https://github.com/dask/dask-examples
- **Technical blog**: wrote articles detailing approaches to my technical work in general. https://andersy005.github.io/
- **Advanced Lane Lines Detection**: A pipeline that uses OpenCV to detect lane lines on the road on a series of individual frames and/or a video stream. https://youtu.be/3NnTZ9NR03k

Technical Skills

Languages	Python, C++, C, Scala, Bash
Frameworks/Libraries	Apache Spark, Numpy, Dask, Xarray Pandas, Scikit-learn, Jupyterlab
Toolchain	UNIX, Git, LaTeX
Cluster Environment	Kubernetes
Infrastructure as service	AWS, EC2, S3, GCP
Build Tools	Travis CI

selected presentations

- **Beyond Matplotlib - Tutorial: Building Interactive Climate Data Visualizations with Bokeh and Friends**: UCAR Software Engineering Assembly. Boulder, CO. April 2018
- **PySpark for "Big" Atmospheric Data Analysis**: American Meteorological Society (AMS) 2018 Conference. Austin, TX. Jan 2018 Recorded Presentation
- **PySpark for "Big" Atmospheric and Oceanic Data Analysis**: National Center for Atmospheric Research. Boulder, CO. Aug 2017 Recorded Presentation