Tom **Augspurger**

Software Engineer

tom.w.augspurger@gmail.com | tomaugspurger.net | TomAugspurger

Summary

Experienced software engineer with a background in distributed computing, data engineering, geospatial data, and managing open-source projects. Active participant in, and organizer of, several open-source and open-science communities, including around pandas, Dask, Pangeo, and STAC. Willing to work hard to solve challenging problems.

Work Experience

Microsoft

Principal Geospatial Engineer — *2020 — 2024*

- Helped design and build the Microsoft Planetary Computer, a geospatial data platform hosting petabytes of data and serving many API requests
- Directly responsible for the geospatial data pipelines and compute platform
- Managed the community of Planetary Computer users
- Helped start geoparquet, a standard for tabular vector geospatial data, and several extensions in the STAC ecosystem
- Managed relationships with partners and vendors working on the Planetary Computer
- Served as team-lead for a squad of three engineers
- Helped design and develop a new Azure service through Private Preview

Anaconda

Software Engineer — *2017 — 2020*

- Maintained several packages in the Scientific Python ecosystem, including pandas and Dask
- Started dask-ml, a library for scalable machine learning
- Worked with users, including the Pangeo community, to solve scaling challenges

Mittera

Data Scientist — *2014 — 2017*

- Performed ad-hoc analysis on customer shopping behavior for several large retailers to estimate customer value
- Built and maintained data pipelines for various internal and customer projects
- Built a client-facing website for monitoring survey data

University of Iowa

Teaching Assistant — *2011 — 2014*

Open Source

Pangeo

Pangeo Steering Council — *2022 — present*

Python Software Foundation

PSF Fellow — *2018 — present*

Pandas

Core developer — *2014 — 2020*

Dask

Core developer — *2017 — 2020*

Various

Maintainer

stac-geoparquet, Zarr, kbatch, dask-geopandas, adlfs, pystac-client

Writing and Teaching

Effective Pandas

Leanpub

Author — *2016*

- A series on writing effective, idiomatic pandas

Clean, effective data analysis with Python

O'Reilly Media

Instructor — *2017*

- Designed and taught a course for O'Reilly's Live Online Training platform

Languages and Tools

Languages

- Experienced in Python and SQL
- Familiar with HTML, CSS, Javascript, LaTeX, and R

Tools and Technologies

- Experienced with the scientific python stack (NumPy, pandas, scikit-learn, xarray)
- Geospatial data management (STAC) and file formats (COG, Zarr, geoparquet, GRIB, HDF5)
- Parallel and distributed computing (Dask, Azure Batch)
- Workflow orchestration engines (Argo, Dagster)
- Web and API frameworks (Django, FastAPI)
- Infrastructure and application deployment (Terraform, Kubernetes, and Helm)
- Monitoring (OpenTelemetry, Azure Monitor, and Microsoft-internal systems)
- Azure (especially Azure Storage, AKS, Event Grid, Cosmos DB, Application Insights, Azure Batch, Azure Functions) and some familiarity with AWS and GCP

Education

University of Iowa

Iowa City, IA

Masters in Economics

2011 — 2014

- Spent three years in the economics PhD program
- Courses in Probability and Stochastic Processes, Optimization Techniques, Econometrics

University of Northern Iowa

Cedar Falls, IA

Bachelors in Economics

2008 — 2011

- 2010 Alumni Scholarship recipient
- Jepson International Economics Essay Contest runner-up (2010) and winner (2011)