Matthew Barber

Southend-on-Sea, Essex, UK 🛛 🔤 quitesimplymatt@gmail.com

FXPFRIFNCF

Data Pipeline Engineer @ Palantir (via Hexegic)

- Built and maintained PvSpark data pipelines for local government agencies across the country as a subcontractor for Palantir.
- Set up monitoring to ensure we know immediately when pipelines are failing, preventing downtime on critical user-facing applications.
- Fixed issues reported by customers, often having to guickly learn code, pipelines and user interfaces written by others to resolve problems under tight SLAs.
- · Monthly on-call rotation to respond to pipeline failures, either fixing them myself or coordinating with other teams.

Software Engineer @ Quansight

July 2021-May 2024

May 2024-

- Helped design API standards as a member of the Python Data APIs consortium, which are now adopted by popular array and dataframe libraries such as NumPy and pandas.
- Built a REST API backend for a client's image detection platform, using FastAPI and various other libraries that modelled, queried and migrated a postgres database.
- Instilled a testing culture in a client's data pipelines team. I encouraged personnel to write tests and CI infrastructure by writing guides, running tutorials and pair programming.
- · Contributed NumPy compatibility layers to PyTorch as part of a contract with Meta. Such layers enable scientific code originally built for NumPy to now take advantage of PyTorch tensors and ops.

PROJECTS

Hypothesis

Popular property-based testing library for Python that I help maintain.

- Contributed array API tools that are now used in the test suites of libraries such as NumPy and PyTorch.
- · Maintained modules used to test numerical/scientific code generally.

dataframe-interchange-tests & array-api-tests

Compliance test suites for the dataframe interchange protocol and array API standards respectively.

- · Employed property-based testing to generate many varied and interesting test cases to say with a high degree of confidence whether an implemented API is compliant or not.
- · Architected library-agnostic tests so that the test suites can run against any implementation of a standard.
- Projects like NumPy and PyArrow have found dozens of bugs in their main libraries just by testing their API compatibility layers.

EDUCATION

Aston University

1st Honours BSc Computer Science

FIND ME ONLINE

- github.com/honno
- blog.honno.dev
- in linkedin.com/in/honno

OPEN SOURCE CONTRIBUTIONS NumPy PyTorch CuPy JAX pandas PyArrow polars Vaex cuDF modin **OpenBLAS** Hypothesis (maintainer) pandera Rich LANGUAGES Python Java C# SQL JavaScript Bash Lisp TOOLS pytest git **GitHub Actions** Jupyter Notebook Linux