Michael Willy, Data Engineer

michaelwilly.com | linkedin.com/in/michaelwilly/ | github.com/chriswilly

- I hold six years of experience developing data projects across research teams and compute platforms.
- I automate Python & SQL to deliver scientific data to global research teams. I use SQL to dimension tables, transform reports, and perform application transactions.
- I developed numerical models on high-dimensional image and text using SQL databases. Interactive results are published on Jupyter Notebooks or Power BI dashboards created with development user feedback.

Tools

Python (Polars SQL, Pandas, NumPy, Scikit-Learn, SQLAlchemy, Dash Plotly), Git, Power BI, LaTeX
SQL (Oracle, MS SQL, SQLite), Amazon AWS S3, Bash & PowerShell, Cron, Airflow, Docker, Markdown

Experience

Data Engineer, Research Chemical Company March 2023 – Present
- I own the operation and data automation of the corporate R&D Electronic Lab Notebook (ELN).
- I am responsible for automated monitoring and data governance on my company's electronic lab notebook. I routinely join separate data sets from databases and APIs, and I introduced git versioning, automated linting, and code review to the opertaion.
- I developed an API pdf render and file downloader using Python HTTP requests and SQL ad hoc queries to export records during a divestiture. This saved the company $50k from a vendor estimate and delivered in three weeks vs months.
- I built automated Python Polars csv reports with LDAP and SQL queries using Linux virtual machines. These jobs include migration document accounting, nightly update validation, and data sanitization.

Data Science & Engineer, Research Chemical Company September 2020 - March 2023
- I developed Python & SQL pipelines to extract lab documents to transform into an Oracle or MS SQL database model.
- I developed a Fast Fourier Transform (FFT) module & plotting pipeline command line CLI tool and Jupyter Notebook interface.
- I built a Dash Plotly and Jupyter Notebooks, and Microsoft Power BI to create a process visualization valued by the tool vendor as a $300k estimate.

Project Engineer, Research Chemical Company September 2018 - September 2020
- I automated time series and part wise database reports using SQL & Python on virtual machines to publish measured process values on a pilot manufacturing line. The product was used in semiconduictor fabrication research and had multivariate performance regression on manufacturing inputs.
- I designed instrument and user interfaces and the pilot facility's data model using MS SQL Server.
- I managed contract equipment installation and start up for nine pilot machines costing $800k capital over two years.

Research Engineer, Research Chemical Company September 2014 - September 2018
- I scaled-up organic light emitting diode (OLED) purification and programmed robotic manufacturing.

Work Projects

- Fast Fourier Transform Numpy & SciPy signal processing automation with Plotly dashboard, JSON API, and runtime SQLite database November 2022 - June 2023
- Scikit-learn Graph Euclidian distance machine learning model to classify handwritten digits February 2024

Education

Master of Science Applied & Computational Mathematics
University of Washington, Seattle, WA December 2023
Bachlor of Science Chemical Engineering, Minor in Mathematics
Drexel University, Philadelphia, PA June 2011

Date: May 8, 2024.