Michael Willy, Data Engineer

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

- I hold over six years of developer experience in data projects serving business and research stakeholders.
- I automate Python & SQL data jobs to deliver relevant information to global research teams. I develop SQL to dimension tables and analytics reports. I use Python for process orchestration and visualization of complex relations and data transformations.
- I developed numerical models and datasets on high-dimensional image and text databases. Interactive results are published using Jupyter Notebooks and Power BI dashboards created alongside user feedback.

SKILLS

Python (Polars SQL, Pandas, NumPy, Scikit-Learn, SQLAlchemy, Dash Plotly), Git, Power BI, LaTeX
SQL (Oracle, MS SQL, SQLite), Amazon AWS S3, Bash & PowerShell, Airflow, Cron, 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. I develop with git versioning, automated linting and testing, and code reviews.
- I developed a rendered document API caller using Python HTTP requests and SQL queries to export records for an intellectual property divestiture. This saved the company $50 to 70k from a vendor estimate and delivered in three weeks vs. months spent on a project.
- I built automated Python Polars csv reports with LDAP and SQL queries. These jobs include access reports, automated rules update validation, and environment data sanitization.

Data Science & Engineer, Research Chemical Company   September 2020 - March 2023
- I developed Python & SQL pipelines to extract lab documents, perform numerical harmonic analysis and fluuid dynamics calculations, then transform to Oracle or MS SQL database models using Pandas and SQLAlchemy ETL tools.
- I built a Dash Plotly animation and Jupyter Notebooks interface, and Microsoft Power BI dashboards 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 a time series database jobs where Python & SQL summarized manufacturing process data.
- I managed contract equipment installation and start up for nine pilot machines valued at $800k capital spend in two years.

Research Engineer, Research Chemical Company   September 2014 - September 2018
- I programmed robotic spincoaters and proportional-integral-derivative controllers (PID) for organic light emitting diode (OLED) fabrication.

WORK PROJECTS

- Application project-level access review for 1,000 active users and 1,000 projects developed for data governance. Automated compliance system uses Python orchestration of Oracle SQL run by cron jobs. Features include LDAP-based tree reporting, automated administrative nightly updates, and matplotlib relation mapping. SQL update validation using Linux mail utils.   May - September 2024
- Fast Fourier Transform Numpy & SciPy digital signal processing automation published to a Plotly dashboard and SQLite runtime database.   November 2022 - June 2023

EDUCATION

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