

DAVID VIGH

(315) 222 - 3004 | DAVID@DAVIDVIGH.COM | DAVIDVIGH.COM | GITHUB.COM/DVIGH8

EXPERIENCE

Amgen - Deerfield, IL

Sr. Machine Learning Engineer

Jan 2025 - Present

- Optimized ML infrastructure and model training pipeline through parallel processing and code refactoring, reducing training time by 75% and infrastructure costs by 50%
- Led production deployment of a transformer-based ML pipeline that automatically generates and ranks qualified sales leads weekly, enabling sales teams to prioritize high-potential opportunities and reduce time-to-contact

Sr. Data Engineer

Oct 2023 - Dec 2024

- Developed and maintained automated ETL scripts to process the team's main data sources from various vendors totalling 200+ billion rows over each month. Published final datasets in Google Cloud, Databricks, and Azure SQL
- Administered working environments for data scientists, including Linux servers and Google Dataproc, ensuring seamless and efficient workflows. Made continual improvements to workflow, processes, utilities, and automated jobs
- Deployed automatic workflows for regularly recurring data pipelines reducing time to insights from weeks to days
- Assigned tasks and provided thorough feedback to direct reports including offshore consultants

Horizon Therapeutics (acquired by Amgen) - Deerfield, IL

Sr. Data Engineer

Jul 2022 - Oct 2023

- Implemented and advocated for Git adoption among our team, streamlining collaboration and codebase management
- Achieved a smooth transition to an upgraded data format, evidenced by uninterrupted operations through comprehensive stakeholder engagement
- Enhanced the scope of data by incorporating new claims vendors, leading to a significant increase in insights across all vendor areas
- Guided and supervised a summer intern, leading to the successful development of an internal doctor lookup tool, enhancing internal resource accessibility

Data Engineer II

Jul 2021 - Jul 2022

- Increased operation ability by increasing cloud footprint across multiple cloud vendors
- Continue MDM project development with stakeholders across analytics departments
- Managing and improving the scalability and capabilities of the data science platform
- Improving ETL pipelines with new features and data sources

Data Engineer

Jan 2020 - Jul 2021

- Decreased processing time by 5x of monthly deliveries of 10B rows of data
- Designed, built, and maintained multiple MDM projects improving organizational clarity of data sets
- Implemented automatic data pulls from vendor websites using web scraping technology due to vendor not having api

Contractor

Aug 2019 - Dec 2019

- Migrated on premises distributed computing environment to GCP to increase platform scalability

Software Developer and Data Engineer - Chicago, IL

Contractor

Jul 2018 - Aug 2019

- Researched and built first stage web scraper
- Built the csv parser for an early stage startup
- Launched several websites for small businesses

EDUCATION

Clarkson University - Masters of Science, Applied Data Science

Clarkson University - Bachelors of Science, Computer Science

SKILLS

Languages: Python, SQL, Go, Scala, Java, BASH, Powershell, HTML/CSS, Javascript, R

Technologies & Frameworks: Pandas, Pyspark, Git, Neo4j, GCP, Azure, AWS, Apache Spark, Snowflake, Databricks, Docker, Delta

Operating Systems: Linux, Windows, Mac