Mark Schott

Old Lyme, CT, USA, 313-908-0388
email: hire.mark.schott@gmail.com
linkedin: linkedin.com/in/meschott
   github: http://github.com/mesnaround

- Data Engineering and Insights expert with a successful track record of high-impact deliveries
- Advanced skills in programming, mathematics, and data visualization
- Reliable, independent, and transparent contributor and team leader
- Passionate about the outdoors, alternative energy tech, FOSS, sustainable land management

Skills and Acquired Experiences:

- **Programming Languages:** Python, SQL, Bash, R, C, Rust
- **Operating Systems:** Linux, Windows (WSL2), Docker Buildx (x86 and ARM)
- **CICD:** Git (Github, Bitbucket, AWS CodeCommit), IaC (AWS CDK, Terraform), Devops (Jenkins, AWS CodePipeline, Git Hooks), Project Management (Jira, Miro)
- **Databases:** MySQL, AWS Glue + Athena, Hudi, InfluxQL, Hive, Impala
- **Python-based:** PySpark, Pandas, Polars, DuckDB, NetworkX, Scikit-Learn
- **Cloud Experience:**
  - *AWS (A lot)*: Lambda, S3, Glue, Athena, Lambda, EC2, EMR, DynamoDB, RDS Aurora, S3
  - *Azure (A little)*: ADLS, azure-cli, Databricks
- **Other Concepts:** SLDC, CICD/Devops/IaC, Multi-cloud setup, Statistical Visualization and Modeling
- **AI/ML:** "Prompt Engineering", Linear Regression, K-Nearest-Neighbors, Random Forest
- **Spoken Languages:** English, French (Advanced Proficiency)

Professional Experience:

Senior Cloud Data Engineer, *EDF Renewables*, Remote, December 2021 – Present

- In collaboration with engineers, SMEs, and stakeholders, end-to-end development of a suite of KPIs (Potential Production, O&M Contracts, Snow, etc) serving over 80 Solar and Storage sites across 6 countries.
- Architected and guided implementation of multiple scalable components that continue to serve the team such as our shared library setup in AWS Lambda and Glue, a data visualization add in InfluxDB+Grafana stack, and a Data ETL to backfill POA Irradiance data.
- Took over sole-engineering of the Solar Python AWS code base from an external team of three developers
- RCA on bugs and data quality issues some alone saving $10,000s per year in processing cost
- Asynchronously working with partners across the organization in Slack and Jira to track and drive features

Data Scientist, *Fiserv*, New Providence, NJ, January 2021 – November 2021

- Dockerized Python app to run in Kubernetes and became a primary developer and code owner
- Created and packaged a Cython interface to an external C library for column-level encryption
- Drove adoption of JIRA and Agile, creating documentation and guiding developers through the process
- Demonstrated methods for local machine development rendering cloud VMs unnecessary
- Worked with internal teams to open firewalls, submit enhancement requests and meet compliance standards

- Developed SQL to visualize big datasets by computing aggregates in Snowflake and plotting in Matplotlib
- Assist in architecture of end-to-end Federated ELT and Data Science platform spanning on-prem and cloud

Senior Data Engineer, *ADP,* Parsippany, NJ, January 2019 – November 2020

- Developer and SME for a business-critical PySpark-based pipeline that drives $100+ million in CARES loans, an international benchmarking product, and a multi-million dollar data monetization business
- Refactored outsourced product to an in-house stack using PySpark on EMR, Glue, and S3, greatly improving performance and flexibility while reducing data refresh time for clients by over two weeks
- Spearheaded the implementation of a time-series outlier detection tool (Python, Scala) that checks metrics while processing and has saved many hours of processing, research, and communications
- Defined data-anonymization standards for EU compliance to allow global benchmarking product roll-out
- Communicated with external clients and internal consumers to answer data product specific questions

Data Engineer, *ADP,* Roseland, NJ, September 2017 – January 2019

- Led comprehensive data quality analysis in Bash, Python, and R using Impala, matplotlib, and ARIMA to visualize time-series data quality, compute monitoring thresholds, and calculate correction factors
- Developed Hive and Spark-SQL transformations for an client product and interfaced with file delivery platform using cron-scheduled Bash scripts which was a key in automating for production
- Revamped an in-house Python-based packaging, deployment, and execution tool adding flexibility which allowed further adoption across our organization

Data Science Training, *NYC Data Science Academy,* New York, NY, April 2017 – July 2017

- Classified next-day movement of bitcoin daily return rates using Support Vector Machines, achieved over 80% specificity and made an R Shiny app to visualize arbitrage opportunities and real-time tweets
- Employed data cleaning, feature engineering, and linear and tree-based modeling techniques to predict real estate prices as part of the Sberbank Housing Kaggle Competition
- Created another R Shiny app to dynamically visualize Consumer Financia qql Protection Bureau complaints

Applications Developer*, Reynolds & Reynolds,* Houston, TX, May 2015 – March 2016
- Designed and developed a process to validate tax data and prevent costly stoppages of monthly billing
- Built a user-customizable management application to boost the efficiency of large scale team management
- Designed an application to handle differences in manufacturer-specific data pertaining to car accessory sales

Experimental Research Intern*, Wayne State University,* Detroit, MI, June 11 – April 2013
- Developed C++ Programs using the High Energy Physics Framework, ROOT, to run simulations, plot histograms, do background subtractions techniques to refine mass measurements for short-lived resonances

Education:

Applied Physics Ph.D. Student, *Rice University,* Houston, TX   Aug.   2013   –
Oct. 2014
BS in Physics (Cum Laude), *Wayne State University,* Detroit, MI   Aug.   2009   –
May 2013