

Mathew Perez

(469) 964-6005 | matperez755@gmail.com | mathewperez.github.io | GitHub: MathewPerez | LinkedIn: mathew-perez

Education

University of Texas at Dallas | MS — *Computer Science*

Richardson, TX | Aug 2019 - Dec 2020

- **Concentration:** Intelligent Systems

University of Texas at Austin | Honors BBA — *Management Information Systems*

Austin, TX | Aug 2016 - Jun 2019

- **Certificate:** Computer Science

Experience

Facebook | Data Engineer

Menlo Park, CA | Dec 2020 – Present

- Implemented and collaborated on data-lineage applications via Java and Scala graph api's to classify over 3k datasets and scheduled daily via Python Dataswarm endpoint
- Helped architect and design multi-tiered data lineage application to traverse across entire data warehouse, conducting privacy lockdown on over 5k Hive datasets, files, ml models, data scrapes, etc.
- Implemented and designed automation bots in Python/Hack to scrape snapshot of Hive Metastore permissions and grant/revoke over 200 dataset access permissions to sensitive data categories asynchronously every 15 minutes, reducing TPM partners' work load
- Helped implement internal privacy tool via React, GraphQL, and MySQL stack to centralize the tracking, alerting, and fixing of thousands of potential privacy violating storage assets
- Implemented and refactored 4 different PHP scripts used across more than 3 teams to automate bulk MySQL db interactions for the internal privacy tool when required by engineering and privacy leadership

Facebook | Data Engineer Intern

Menlo Park, CA | May 2019 – Aug 2019

- Designed, built, and tested Data Swarm pipelines that synthesized over 6 months of data from multiple organizations providing over 10 new specialized and aggregate metrics
- Created custom Jinja template to implement intuitive data dashboard with four bottom line metrics and over 12 graphs
- Worked closely with mobile platform engineers to correctly source more than six client and server side logging data tables for ETL design and implementation
- Formulated three different Presto-generated metrics and graphs with ML Engineers and Product Managers to prove value of adding my created metrics to current ML models' features

University of Texas at Austin | Undergraduate Research Assistant (Dr.Rajiv Garg)

Austin, TX | Feb 2019 – May 2019

- Applied Python NumPy library and csv module to clean and transform over 5 GB of image tensor data to feed into Convolutional Neural Network
- Built Java application through restfb API and login dialog through Facebook JavaScript SDK to automate social media research survey

Skills

Languages Python, SQL, Java, Scala, C/C++, PHP, R, Swift, HTML/CSS

Tools & Tech UNIX/LINUX, Git, Spark, Hive, Hadoop, Presto, DataSwarm, React, NodeJS, Flask

Projects

Austin Environmental Directory Website | Senior Capstone Project — *Team of 2*

February 2019 - May 2019

- Overhauled the MariaDB backend point to decrease web page load times and simplify website maintenance for stakeholders
- Modernized the Directory website by rewriting over 20 PHP and CSS files

Podcast Search and Summary Algorithms | Graduate Research — *Thesis Project*

May 2020 - Nov 2020

- Using Deep Learning NLP techniques in PyTorch on over 13GB of podcast data to create relevant summaries of podcast episodes
- Pairing classical IR methods with Deep Learning in Lucene inverted index to build podcast excerpt search engine