

JOHN PATRICK ROACH

Data engineer with multiple patents and an MS in Statistics, hoping to apply my skills towards big data and tech.

johnpatrickroach.com

(601) 502-7337

github.com/johnpatrickroach

contact@johnpatrickroach.com

SKILLS

- **Code Languages:** Python 3.9.5, Bash/Shell, PHP, osascript, C++, Flutter (Dart), Scala, R, Badger
- **AWS:** EC2, S3, Redshift, Athena, RDS, Lambda, Kinesis, ELB, EMR, Neptune, CloudFront, VPC, SNS
- **Web Scraping:** Multiprocess, asynchronous, memory efficient web scraping, API engineering
- **Data Science:** Machine Learning, Classification, Regression, KNN, Neural Networks, NLP
- **Cloud Services:** AWS, Google Cloud Platform, Azure, Snowflake
- **Databases:** dgraph, Redshift, NoSQL, MySQL, Postgresql, Redis, Elasticsearch, MongoDB, Apache Druid
- **Web Development:** React, Next.js, HTML, CSS, Javascript, nginx, Django, Flask
- **Tools:** Spark, Pyspark, Plotly, Dash, Numpy, Pandas, Scikit, Keras, Pytorch, strace, Algolia, Postman, Burp

PATENTS

Oct. 22, 2019	INDIVIDUALIZED CONNECTIVITY BASED REQUEST HANDLING	10455008
Oct. 01, 2019	LOW-LATENCY HIGH-THROUGHPUT SCALABLE DATA CACHING	10432706
Oct. 01, 2019	GEO-PARTITIONED DATA CACHING	10432737
Jun. 05, 2018	DYNAMIC ADVERTISING ROUTING	9992121

EXPERIENCE

04/21 to 06/21	Consumer Edge <i>Senior Software Developer</i>	New York, NY
	<ul style="list-style-type: none">• Used their proprietary Savvr (PHP) web-scraping system to collect various data from 40+ publicly traded company sites, such as eBay, Carmax, etc...• Synthesized raw scraped data into SQL database tables for higher level KPI tracking and analysis, used to forecast company performance against Street earnings estimates• Wrote multi-processed, asynchronous, generator functions in Python to stream 160+ million json responses from public eBay API endpoints and stored the data in Google Cloud Storage for efficient analysis using BigQuery	
12/20 to 04/21	FanDuel <i>Data Engineer</i>	New York, NY
	<ul style="list-style-type: none">• Worked with a team of data engineers to improve and maintain data pipelines for a variety of FanDuel products• Updated Redshift tables for more efficient data queries and processing	

- Refactored Airflow configurations and DAG structures to improve data pipelines and ETL processes

02/20 to 12/20

BetterOptions.io

New York, NY

Co-Founder and Software Development

- Developed systematic trading screener, execution and portfolio manager for non-linear instruments with an integrated API
- Developed a multi-process, asynchronous web-scraping framework in Python3.8 to collect financial data
- Built and tested automated personal stock trading systems on top of Robinhood API
- Designed and developed a modern database using Dgraph, a cloud-based, horizontally scalable and distributed GraphQL database with a graph backend.
- Designed and developed a modern frontend using Flutter, Google's UI toolkit for crafting beautiful, natively compiled applications for mobile, web, and desktop from a single codebase.
- Built a Python wrapper for text-mining and sentiment analysis through the StockTwits API

02/20 to 12/20

Deep Root Records

New York, NY

Contract Web Developer

- Developed the full stack of a web site for the Licensing division of Deep Root Records (deeprootlicensing.com)
- Example showcase of the backend: youtu.be/gwI07GDo3eE
- Designed and developed custom, native search functionality for searching songs by artist, genres, themes, BPM, and key using Algolia and Next.js (a wrapper for React): browse.deeprootlicensing.com
- Used S3 for hosting the audio files and related artwork
- Designed and developed a native web audio player for users to listen to the tracks directly from their browser, modeling the UI after Spotify's web player
- Designed and developed an internal dashboard and suite of tools for adding, updating, or deleting songs using Retool
- Deployments were managed through git, Github, and Vercel

11/16 to 01/20

EMX Digital

New York, NY

Data Engineer, Software Development

- Worked with EVP of Technology to develop high-frequency, low-latency, distributed online ad exchanges
- \$300k+ in revenue and 50-70 billion transactions (rows of data) per day processed through our data pipelines
- Worked with a team of data scientists to build machine learning models and products on top of our database
- Repurposed obscure telecom software to handle online ad auctions in under 200ms processing time

- Built a custom data streaming service in Bash that ensured zero data loss and sub-second process time at scale
- Design and develop ETL processes and Python automations in AWS environment to migrate campaign data from external sources like S3, into different database storages like AWS Redshift, Apache Druid, Athena, Elasticsearch, Redis.
- Built and maintained a cloud, partitioned, Redis datastore to hold 2.1billion rows of cookie data
- Top-volume client with AWS, actively engaged in product usage negotiations and meetings with AWS leaders

8/13 to 10/16

CPXi

New York, NY

Data Analytics

- Managed millions of dollars worth of native advertising campaigns for various owned and operated websites
- Optimized the cost to volume ratio for maximum profit using custom built regression models
- Generated an average of \$12k profit per day, with some days reaching over \$100k profit

EDUCATION

2011 to 2012

UNIVERSITY OF NEW ORLEANS

New Orleans, LA

Master of Science in Statistics

2006 to 2011

DELTA STATE UNIVERSITY

Cleveland, MS

Bachelor of Science in Mathematics

ADDITIONAL PROJECTS

- Certified in FOREX TRADING SECRETS OF THE PROS WITH AMAZON AWS
- Built social network map, interest map, sentiment analysis and geo-map of Twitter data
- Drummer in the band Social Creatures