Abhinav Gaddipati

7770 Wentworth Drive, Duluth, GA 30097 • 770.715.7240 • Gaddipatia@gmail.com

EDUCATION

B.S. Computer Science + Statistics
(2019 - 2023)
University of Illinois at Urbana–Champaign

MIT MicroMasters
Data Science & Statistics
Jan 2025 – 2026

CITIZENSHIP

U.S.A

CERTIFICATIONS

AWS Certified Machine Learning Specialty - 2023
AWS Certified Cloud Practitioner - 2019
Machine Learning Certificate from Stanford - 2020
Google Associate Cloud Engineer - 2021

WORK EXPERIENCE

Data Engineer – Greenphire LLC. (Thoma Bravo portfolio company) Atlanta, GA (September 2023-Present)

Responsible for designing, building, and maintaining scalable data pipelines, machine learning models, cloud infrastructure, APIs, DevOps practices, data quality frameworks, and visualization solutions to support clinical trial operations across the enterprise.

- **Designed and delivered a Patient Spend Budgeting solution** on AWS to accurately forecast clinical trial participant expenses, removing financial barriers and improving recruitment, retention, and overall trial integrity.
    - Ingested and processed data from multiple financial systems.
    - Cleaned, cataloged, and aggregated data using PySpark in AWS Glue following data lake patterns.
    - Performed anomaly detection with feedback loops to auto-correct data issues.
    - Exposed curated datasets via FastAPI/API Gateway and deployed infrastructure using Terraform and GitHub Actions.
- **Built and deployed time series forecasting models like DeepAR using AWS SageMaker** to predict clinical trial budgeting needs. These models were trained on historical spend data and incorporated seasonality, trial phase indicators, and participant count to drive accurate projections.
- **Architected and implemented a company-wide reporting framework** using FastAPI (backend) and Angular (frontend), integrated with Ping for SSO-based authentication.
    - Embedded Qlik charts for rich, interactive data visualizations.
    - Deployed as a Kubernetes service and seamlessly integrated into Greenphire's main application as a secure, self-service reporting platform for Greenphire's customers
- **Translated legacy SQL-based ETL pipelines into PySpark jobs** on AWS Glue, significantly improving scalability and reducing run-times for large data sets across multiple environments.
- **Developed a reusable Terraform module for AWS Glue**, enabling any data engineer in the company to rapidly deploy Glue jobs, crawlers, triggers, and workflows across dev, QA, and prod environments with minimal configuration.
- **Implemented company-wide CI/CD pipelines** using **GitHub Actions** to automate testing, infrastructure provisioning, application builds, and deployment workflows for all cloud, API, and data services.
- **Built robust data quality validation frameworks**, alerting, and self-healing workflows using AWS services including Lambda, CloudWatch, S3, EventBridge, and Athena.

SKILLS

Python • C • C++ • OCAML • JavaScript • System Verilog • React • Angular • Django • Scala • Spark • Octave • MongoDB • SQL • Kafka • HTML • CSS • AWS • PSQL • Sagemaker • Glue • Lambda • S3 • API Gateway • CloudWatch, Athena • Systems Manager • EventBridge • RDS • ECR • EKS • EC2 • CloudTrail • VPC • Secrets Manager • Terraform • PySpark • GitHub Actions • Python • Kubernetes • Helm • Flux

- **Led the containerization and deployment of applications** using Kubernetes, Helm charts, and FluxCD for GitOps, ensuring consistency and scalability across microservices.
- **Proficient with AWS Glue and PySpark**, including complex transformations, dynamic frames, performance tuning, and advanced partitioning strategies for large-scale data processing.
- **Technologies:** AWS Glue, Lambda, S3, Athena, API Gateway, EventBridge, RDS, EC2, EKS, ECR, Systems Manager, Secrets Manager, CloudTrail, VPC, CloudWatch, PySpark, Python, Terraform, GitHub Actions, FastAPI, Angular, Kubernetes, Helm, Flux, Qlik.

- **Recognition:** Nominated for the **Phirestarter Award**, recognizing exceptional contributions and exceeding expectations in impact and innovation.

Software Engineer - Checkbook.io, San Mateo, CA (May 2022 – December 2022)

Checkbook is a leading payment platform enabling businesses to send and receive funds via email through a single API call—offering instant payments with the affordability of ACH and eliminating the need for recipients to sign up or share banking details.

- Enhanced the internal content management system by optimizing performance using **Gatsby** and **Netlify**, improving site load speed and content deployment workflow.
- Designed and implemented a robust **address verification system** for new users using **Angular** and **Python**, ensuring improved data accuracy and compliance.
- Built and released new features to streamline digital check creation, resulting in a more intuitive and efficient user experience.
- Upgraded the application frontend to **React 18.10**, improving performance and enabling modern features and libraries.
- Diagnosed and resolved critical UI/UX bugs, boosting platform reliability and reducing support tickets.

- **Technologies**: **AWS, React, Angular, Python, Django, Gatsby, Netlify, PostgreSQL**

Machine Learning Engineer Intern- SemiCab, Atlanta (May 2021-August 2021)

SemiCab is a collaborative transportation platform that leverages predictive analytics to optimize millions of freight loads and thousands of trucks across the U.S.

- Developed and evaluated multiple machine learning models—including **Moving Average, SES, Holt-Winters, SARIMA**, and **DeepAR**—to forecast truck demand at specific locations and times.
- Identified **DeepAR** as the most effective model based on RMSE and final loss metrics, significantly outperforming traditional time series methods.
- Tuned hyperparameters to improve model performance, reducing RMSE and final loss through iterative experimentation.
- Deployed the DeepAR model on **AWS SageMaker** and exposed it via a scalable API endpoint.
- Built an integrated **Flask web application** connected to the SageMaker endpoint, allowing business analysts to test forecasting models through a user-friendly interface.

- **Technologies**: **Python, Pandas, AWS SageMaker, S3, Flask, HTML, CSS**

Computer Science Summer Institute - Google, Atlanta (July 2019-August 2019)

- Developed a full-stack web application that aggregates real-time inventory data from multiple merchants using **Google-provided APIs**.
- Designed and implemented a dynamic interface that allows users to compare product details such as **price, specifications, and availability**, enhancing the shopping experience.
- Collaborated in a team environment to follow best practices in web development and software design.

- **Technologies**: **HTML, CSS, JavaScript, Python, SQL**

Big Data Intern - Cardlytics, Atlanta (May 2018 - August 2018)

Cardlytics is a purchase intelligence platform that leverages real-time consumer transaction data to deliver personalized offers and insights.

- Developed scalable data pipelines to process real-time consumer purchase data and feed it into Cardlytics' **Purchase Intelligence Platform** for targeted marketing campaigns.
- Built and optimized **batch and streaming queries** using **Kafka** and **Scala/Spark**, enabling low-latency data ingestion and transformation.
- Designed and implemented **ETL workflows** with **Spark/Scala**, transforming raw transactional data into actionable insights.
- Deployed and configured a **Hadoop cluster** to support large-scale data storage and distributed processing.
- Engineered data loading processes from **Hadoop** into **Vertica** to support advanced analytics and reporting.
- **Technologies**: **Kafka, Scala, Spark, Hadoop, Vertica, AWS, SQL, Python, S3**