

Pranav Modh

✉ modhpranav21@gmail.com 📞 +1 8573859511 📍 United States 🌐 modhpranav/profile in in/modhpranav

Skills	
Programming Languages (Python, Java, R, Javascript, Unix/Linux)	Data Engineering Platforms (Snowflake, Databricks, Palantir Foundry)
Big Data Technologies (Apache Spark (including PySpark), Hadoop, Kafka, HDFS, MapReduce)	Frameworks (Rest Framework, Flask, FastAPI)
Cloud Services (AWS (EC2, Lambda, S3, Redshift, SQS) GCP (cloud functions, Topics, Big Query))	
Databases (DynamoDB, PostgreSQL, MySQL, MongoDB)	Containerizing Tools (Docker, Docker Compose, Kubernetes)
Python Libraries For Data Processing (Spacy, NumPy, SciPy, Pandas, Matplotlib, Elasticsearch, PySpark)	

Professional Experience (3.5+ years)	
MilliporeSigma (Merck KGaA), Boston, United States, Data Engineer Intern ⓘ	05-2023 – 12-2023
<ul style="list-style-type: none">Optimized data processing and analysis in the semiconductor industry using Palantir Foundry, Python, and Apache Spark.Integrated a supply chain data pipeline with AWS services (Lambda, S3, Redshift) and Kafka, enhancing logistics and operational efficiency.Employed enhanced algorithm for demand forecasting, resulting in a 20% reduction in overstock costs.	
Simform Solutions, India, Software Engineer ⓘ	12-2021 – 06-2022
<ul style="list-style-type: none">Spearheaded the end-to-end development of the company's first data pipeline, harnessing Facebook Graph API, AWS S3, Apache Kafka, Hadoop HDFS, and Airflow, to streamline data accumulation, storage, and processing.	
Tntra IO, India, Associate Software Engineer ⓘ	08-2019 – 12-2021
<ul style="list-style-type: none">Advanced from Junior Python Developer, leading the development of microservices for diverse data processing, utilizing technologies like Python, Flask, and MongoDB.Engineered a sophisticated web scraping module to extract and structure data from various online sources, significantly enhancing the data acquisition process.Orchestrated client communications and task management, streamlining project execution and ensuring efficient resolution of technical issues.Developed a Project Management Tool using Django, and integrating with PostgreSQL for efficient data management, coupled with Docker-Compose for containerization, leading to a 15% reduction in redundant module construction.	

Recent Projects	
Supply Chain Optimizer For SemiConductors (EMD Digital - Merck Group)	
<ul style="list-style-type: none">Utilized Palantir Foundry's in-house applications for efficient large-scale data processing, contributing to a 30% reduction in interdependencies and a 25% increase in time efficiency for route optimization.Implemented geolocation-based batch location optimization using AWS Lambda, resulting in a 15% reduction in supply costs.Developed real-time reporting features, leveraging Palantir Foundry and AWS technologies, leading to 20% faster decision-making by enabling immediate access to actionable data insights.	
Covid Hospitalization And AQI (Report For MilliporeSigma)	
<ul style="list-style-type: none">Orchestrated an automated data pipeline using Apache Airflow, ensuring efficient and consistent data flow from public health databases into the Hadoop Distributed File System (HDFS) for scalable storage.Leveraged Apache Spark and MapReduce (MRJobs) for comprehensive data analysis, identifying correlations between air pollution and COVID-19 hospitalization risk, while efficiently managing large-scale datasets.Utilized advanced visualization tools (D3.js) to present complex data insights, enabling stakeholders to interactively explore trends and patterns in air quality and hospitalization data.	
Image Processing And Exploration	
<ul style="list-style-type: none">Streamlined preprocessing of 500K+ user-uploaded documents in diverse formats (EML, PDF, RTE, DOC, DOCX) using Apache Kafka for efficient real-time data streaming.Leveraged Python and Informatica for format conversion and image preparation, followed by hashing algorithms for precise categorization.Used Docker for efficient deployment on AWS EC2, enhancing scalability, and fine-tuned a Faster R-CNN model on SageMaker for high-accuracy format recognition.	
Freight Management (Tntra.io)	
<ul style="list-style-type: none">Played pivotal role by contributing over 70% to project's development, establishing it as industry standard platform for all freight businesses.Crafted custom web scraping scripts using Python's Pypeteer for data extraction from 15 distinct international freight websites. Transformed and ingested data into a database for standardized processing.Automated data pipeline on AWS using SQS, Lambda, S3, and Redis. User-uploaded XML files trigger Celery and Redis to enqueue container numbers in AWS SQS. Lambda functions scrape, preprocess, and store data in S3, with email notifications for user access via a Django app.	

Education (Expected Graduation: May 2024)	
MS Computer Science (GPA: 3.7), University of Massachusetts, Boston, United States ⓘ	09-2022 – present
B.E. in Electronics and Communication, Gujarat Technical University, Gujarat, India ⓘ	07-2015 – 05-2019