

Rahul

6+ Years | Data Engineer

Summary:

- **6+ years of hands-on experience** as a Data Engineer and AI/ML Developer, designing and implementing **datadriven solutions** using **Google Cloud Platform (GCP)**.
- Proficient in **GCP services** such as **Dataflow, Cloud Composer, Cloud Functions**, and **GCS Buckets** for building scalable data pipelines, orchestrating workflows, and automating backend processes.
- Skilled in analyzing large datasets, generating **data visualizations**, and delivering **insightful reports** to support **data-driven decision-making** across business units.
- Experienced in developing **end-to-end architectures** for machine learning workflows and big data processing using **cloud-native technologies**.
- Strong in **data modeling, ETL/ELT design**, and applying **AI/ML algorithms** to solve real-world business problems. • Excellent **analytical and problem-solving skills**, with a focus on building **efficient, reliable, and secure** data infrastructure.
- A collaborative and detail-oriented professional with a proven ability to work in **cross-functional teams**, driving performance improvements and business impact.

Technical Skills:

Programming Language	Python, R, SQL, JavaScript, TypeScript, Shell/Bash
Data Visualization	Tableau, Power BI, Looker, Google Data Studio, Plotly / Dash (Python)
Frameworks	TensorFlow, PyTorch, Langchain Framework, AI Frameworks, Flask / FastAPI
NLP & CV	Natural Language Processing (NLP), Computer Vision, SpaCy, NLTK
Cloud Platforms	AWS, Azure, GCP, BigQuery, Dataflow, Cloud Functions Cloud Storage (GCS)
Containerization	Docker, Kubernetes, Google Kubernetes Engine (GKE)
Databases	SQL, NoSQL, Vector Database, Spark, Hadoop, PostgreSQL, BigQuery, MongoDB, Redis
DevOps & MLOps	MLOps, CI/CD, Git ,GitHub ,GitLab, Jenkins, GitHub Actions ,Cloud Build
Other Technologies	Explainability, AI Ethics, Bias Detection, Great Expectations, JSON / YAML, Google Analytics
Generative AI	ChatGPT OpenAI APIs, LLM, SFT, RLHF, AI Model Training
Tools	DVC, MLFlow, Docker, Kubernetes, Spark, Hadoop, SageMaker, MLFlow, or Kubeflow, NLP, JIRA, Confluence, Notion, ClickUp
Database Language	Advanced SQL, BigQuery SQL, PL/pgSQL
ntainerization & Orchestration	Kubernetes, Docker
MLOps & Model Deployment	CI/CD, MLFlow, Azure ML, GIT, FastAPI, Model Deployment

Debugging & Process Improvement	Issue Resolution, System Optimization
Monitoring & Observability:	Grafana, Prometheus, OpenTelemetry (instrumentation for pipeline latency and model failure tracking)
Exploratory Data Analysis (EDA)	Data Cleaning, Feature Engineering
Team & Stakeholder Management	Team Management, Stakeholder Collaboration, Coordination

Project Experience:

Project 1: AI/ML Initiatives for Market Research, Transportation & Retail

Role: AI/ML & Data Engineer

Duration: 1 year 8 months

Team Size: 7

Description:

Led AI/ML initiatives across APAC, NAM, and LATAM regions, delivering predictive modeling solutions for pricing, customer behavior analysis, and logistics automation across diverse domains including government and retail.

Responsibilities:

- Developed end-to-end AI/ML pipelines focused on price prediction models for the logistics sector using **Python** and **ML frameworks**.
- Architected and deployed solutions on **GCP**, leveraging **Dataflow**, **Cloud Composer**, and **Cloud Functions** for orchestration and automation.
- Integrated OpenTelemetry SDK in Python pipelines and exported performance metrics to Prometheus and Grafana for real-time monitoring.
- Developed anomaly detection models to flag unusual latency and error rates in batch and streaming pipelines.
- Utilized **GCS Buckets** for scalable and secure data storage.
- Designed and implemented dynamic pricing models for **Less Than Truckload (LTL)** shipments.
- Delivered models for customer affinity, churn prediction, and Customer Lifetime Value (CLTV) for a UAE government client.
- Built real-time **shipment pricing simulator** using **Dash**.
- Automated marketing dashboards using **Power BI** for real-time campaign insights.
- Developed an internal **HRMS tool** to address employee attrition.
- Led a cross-functional team of 7 data scientists and analysts, overseeing the full ML lifecycle and project architecture.

Technical Stack: Python, Machine Learning, AWS, Azure, Elasticsearch, Power BI, Dash, SQL, GCP (Dataflow, Cloud Composer, Cloud Functions, GCS)

Project 2: Retail & Customer Analytics for NAM Markets

Role: AI/ML & Data Engineer

Duration: 1 year 2 month

Team Size: 5

Description:

Delivered personalized targeting, churn prediction, and campaign analytics for major North American retailers like SEG and Tesco, driving data-informed marketing strategies.

Responsibilities:

- Designed and implemented AI/ML models for customer segmentation and hyper-personalized targeting.
- Built and deployed **sentiment analysis models** using **NLP** for refined targeting.
- Created **voice transcript-based product recommendation systems** for medical domains.
- Automated campaign performance reporting using **Power BI** and **VBA**, reducing manual effort.
- Mentored and managed a team of 5 data scientists and interns, supporting delivery and technical growth.
- Integrated GCP services including **Cloud Functions** for automation and **GCS** for storage.
- Contributed to overall **solution architecture** and client engagement from design to delivery.

Technical Stack: Python, NLP, Power BI, VBA, SQL, ML Models, Sentiment Analysis, GCP (Cloud Functions, GCS)

Project 3: Customer Targeting & Delivery Experience Optimization – Amazon DEX IN

Role: AI/ML & Data Engineer

Duration: 1 year 6 months

Team Size: 7

Description:

Partnered with Amazon's DEX IN core team to enhance customer targeting and delivery experiences through machine learning and NLP-based insights.

Responsibilities:

- Migrated existing ML models to **PySpark** for scalable processing within the Hadoop ecosystem.
- Developed churn prediction models and customer segmentation using **clustering techniques**.
- Applied **Apriori algorithm** for frequent product recommendations and **OLS regression** for impact evaluation.
- Integrated OpenTelemetry SDK in Python pipelines and exported performance metrics to Prometheus and Grafana for real-time monitoring.
- Developed anomaly detection models to flag unusual latency and error rates in batch and streaming pipelines.
- Built real-time interactive dashboards with **Power BI** and **Tableau** to support business decisions.
- Applied **NLP** techniques for sentiment classification and review analysis.
- Extracted and processed large datasets using **SQL** from **Amazon Redshift**.
- Engineered robust ML pipelines with monitoring and logging capabilities.
- Contributed to **cloud-native architecture** on **GCP**, integrating **Dataflow** for stream/batch processing and **Cloud Composer** for workflow management.
- Used **Cloud Functions** for event-based triggers and **GCS** for storage pipelines.

Technical Stack: Python, SQL, Redshift, PySpark, Power BI, Tableau, NLP, ML (Logistic Regression, Clustering, Apriori, OLS), GCP (Dataflow, Cloud Composer, Cloud Functions, GCS)

Project 4: Meter Analytics & Customer Churn Prediction

Role: AI/ML & Data Engineer

Duration: 2 Year **Team**

Size: 5

Description:

Developed advanced analytical and machine learning solutions to monitor meter activity (replacements, installations, and shifts) and predict customer churn, enabling the business to improve retention strategies and profitability through loyalty programs.

Responsibilities:

- Designed and implemented a daily reporting system to monitor key meter activities, including installations, shifts, and replacements.
- Created interactive dashboards using **Excel** for productivity and operational insights.
- Developed **customer churn prediction models** using **Python** and **R**, aiding in the development of targeted loyalty and retention strategies.
- Utilized statistical modeling techniques to improve the accuracy and reliability of churn predictions.
- Collected, cleaned, and prepared training datasets for machine learning model development and evaluation.
- Contributed to project planning and execution in collaboration with data scientists and business stakeholders.

Technical Stack: Python, R, Excel, Machine Learning, Data Visualization

