

RIDHI THAKUR

Bay Area - 94085 | (650) 918-2990 | ridhithakur826@gmail.com | <http://www.linkedin.com/in/ridhithakur/> | www.ridhithaakur.com

OBJECTIVE

Results-driven pre-sales professional with 5+ years guiding Fortune 500 clients through AI and data transformations, delivering tailored demos, drafting clear scopes of work, and securing technical wins across healthcare, finance, logistics, and infrastructure sectors. I excel at leading cross-functional workshops to translate complex requirements into actionable solution plans, leveraging GenAI and LLM frameworks to prototype innovative enterprise use cases, and building strong customer relationships that drive solution adoption and maximize ROI.

CERTIFICATIONS

- Data engineer Associate (**Databricks**)
- AWS Certified Solutions Architect Associate (**Amazon**)
- Certified AI associate (**Salesforce**)
- Cloud Data Engineer (**Google**)
- AWS Generative AI Applications (**Amazon**)
- Software Engineering (**JP Morgan**)

EXPERIENCE

Google @Compass, CA, USA | Role : Data Engineer / Analyst

Apr 2024- Present

- Partnered with product and engineering to convert Financial Engineering objectives into actionable data requirements and SLAs.
- Authored and optimized advanced **SQL** models in **BigQuery** for revenue, churn, and payroll metrics; cut query cost 30 %.
- Launched self-service **Tableau** and **Looker** dashboards adopted by 150+ users for authentication and subscription analytics.
- Built automated data-validation pipelines (Python, Airflow, dbt) that trigger real-time alerts via Slack and email, reducing incident MTTR by 70 %.
- Unified Google Analytics and Heap events with internal logs to create a privacy-compliant customer 360° dataset in **Snowflake**.
- Decreased Spark + Airflow latency by 75 % after migrating compute to Databricks and optimizing partitioning strategy.
- Drove 10+ PoCs from pilot to production, directly informing two roadmap features related to identity analytics.

MedImpact, CA, USA | Role: Jr. Data Architect

Sep 2023 - Feb 2024

- Led a next-gen digital transformation by digitizing legacy health questionnaires and pharmacy-benefit forms—capturing 100 M+ daily events via REST APIs and microservices
- Was a part of the OCR-digitizing 500 K handwritten claims with Python/Tesseract/OpenCV, saving \$250 K and cutting manual entry time 80 %.
- Architected governed data pipelines using **Snowflake**, **Databricks**, **Spark**, **Hadoop**, **Hive**, and **dbt** under strict HIPAA compliance—slashing ad-hoc query latency 60%, enabling anomaly detection with Python/sklearn, and achieving audit readiness ahead of schedule.
- Dockerized microservices and built CI/CD pipelines with **Git** and **Jenkins**, reducing release cycles from weekly to daily.
- Delivered real-time self-service dashboards in **Tableau** and **Power BI** for ~200 K patients and 200+ clinicians, boosting analytics engagement 35% and driving proactive care decisions.
- Conducted root-cause analyses that lifted data-trust scores from 88% to 98% and cut escalations by one-third, while partnering with finance to realize \$600 K in incremental revenue within six months.

SDSU Research Lab, CA, USA | Role: Data Analytics Research Assistant

Aug 2021- Jul 2023

- Developed and tested memory-efficient AI models for port operations at Port San Diego, improving digital workflow efficiency by 95%.
- Built financial analytics dashboards using **Bloomberg** data; improved equity prediction metrics for academic research.
- Co-authored research papers on AI, cybersecurity, and information science methodologies.
- Led a team that built an ML-driven maritime analytics system which helped streamline berth allocation using historical traffic trends.

TideWorks Technology, WA USA | Role: Data Engineer – Intern

Jun 2022 - Aug 2022

- Streamed IoT signals from 100+ ports via **Kafka** into **BigQuery**, powering near-real-time logistics dashboards.
- Built **Spark** ML models achieving 87 % accuracy in predicting vessel turnaround times, surfaced in **Looker** reports.
- Authored dbt tests and Python validation scripts to enforce schema and quality checks, catching data issues early.
- Partnered with product managers to define analytics OKRs, influencing two new SaaS features for port clients.
- Documented data lineage and created a metadata dictionary in **Confluence**, reducing onboarding time for new engineers by 30 %.

Eagle Infra Pvt. Ltd, IND | Role: Data Engineer

Dec 2018 - Jul 2021

- Architected SmartCity data lake with **Fivetran**, **Spark**, and **Snowflake** to ingest sensor, traffic, and utility feeds.
- Developed anomaly-detection and forecasting models (ARIMA, TensorFlow), reducing service disruptions by 35 %.
- Built interactive **Power BI** and **Tableau** dashboards for city planners, informing infrastructure maintenance and resource allocation.
- Implemented **SQL** validation tests and governance workflows, cutting data-quality escalations by 40 %.
- Led workshops with municipal stakeholders to refine metrics and prioritize analytics features, ensuring policy relevance.
- Automated ETL pipelines with **Airflow**, reducing runtime by 60 % and improving data freshness by 2 hours.

PROJECTS

Intelligent Traffic Management and Violation Detection System (Python, SQL)

- Developed a YOLO real-time object detection system with Darknet to identify vehicles like Cars, Buses, Motorbikes, Trucks, and Autos.
- Moderated traffic violations, including red light jumping and helmet-less motorbikes, automating violator image capture and number plate extraction.

Intelligent System for Weather and Congestion Probing (Python, SQL, Tableau)

- Implemented advanced forecasting models including ARIMA and RNN-LSTM to assess accuracy of weather predictions.
- Utilized KNN algorithm to estimate commute time by accounting historical data and incorporating next day's temperature information from ARIMA.

Secure and Privacy Preserving Keyword Search Retrieval Over Cloud Data (HTML, JAVA, MYSQL, AWS)

- Performed a cloud technology model to protect important records of users by searchable cryptographic techniques to improve security of records that are saved in the cloud.
- SPHF & SHA-1 algorithm was used to perform hashing technique, the Code project was written in Java and had SQL database.

E-Commerce Geospatial analysis for European market (Runner-up project in MSIS department of SDSU – R, SQL)

- Built a mock e-commerce web-application for order placement and purchasing, utilizing Python for coding and Power BI.
- Analyzed the European E-commerce market during the Covid-19 pandemic using Apriori algorithms, Association rules, and generated Heat maps.

EDUCATION

San Diego State University, San Diego, United States

Aug 2021 - May 2023

Master of Science in Information Systems (Data Science and Statistics specialization)

Jawaharlal Nehru Technological University, Hyderabad, India

Jul 2016 - May 2020

Bachelor of Technology in Information Technology (Statistics Modelling and Probability Specialisation - 3.8/4)

TECHNICAL SKILLS

- **Languages & OS:** Python, R, SQL, Java, Scala, TypeScript, JavaScript, C/C++, Perl, PHP, HTML, CSS, JSON, XML; Linux
- **Frameworks & APIs:** Node.js, React, Angular, GraphQL, RESTful APIs, Presto
- **Data Engineering & Orchestration:** dbt, Airflow, Terraform, Docker, Kubernetes
- **BI & Visualization:** Tableau, Power BI, Looker, ThoughtSpot, MS Excel
- **Cloud & DevOps:** AWS (S3, EC2, Lambda, Redshift, DynamoDB), GCP (Cloud Storage, Pub/Sub, Composer, GKE), Azure Synapse;
- **Jenkins, Git, GitHub Actions, Ansible, Terraform, YAML, JIRA**
- **Machine Learning & Analytics:** Spark MLlib, TensorFlow, PyTorch, Keras, Scikit-learn, NLTK, spaCy, ARIMA, RNN, time series analysis, classification & regression, NLP
- **Monitoring & Logging:** Splunk, Grafana, AWS CloudWatch
- **Methodologies & Presales:** Agile Scrum, stakeholder workshops, reference architecture design, SOW authoring
- **Data Platforms & Databases:** Snowflake, Databricks Delta Lake, BigQuery, MongoDB, PostgreSQL, Teradata, Spanner, SQLite, NoSQL, HBase, Hadoop, Hive, Spark, Kafka, Fivetra