ANKIT MALLYA
Data Engineer

📍 **Mumbai, India**

📞 **+91 9137210394**

💬 **ankit.mallya@gmail.com**

Driven and results-oriented Data Engineer with nearly 4 years of extensive expertise in data engineering, ETL processes, and cloud technologies. Proven track record in implementing robust data pipelines using **AWS (Glue, EMR, Redshift)** and **GCP (BigQuery, Dataflow)**, optimizing data transformations with **Apache Spark** and **PySpark**, and ensuring compliance with data governance standards (GDPR, HIPAA). Skilled in orchestrating workflows with **Apache Airflow**, real-time data streaming with **Kafka**, and leveraging **SQL** and **NoSQL** databases for data modeling and warehousing. Proficient in machine learning, statistical analysis, and automation testing to deliver innovative solutions. Seeking to leverage advanced skills in a dynamic and collaborative environment as a **Product Manager**, **Technology/IT Consultant**, or **Data Engineer**.

CORE COMPETENCIES

Data Engineering & Data Ingestion

ETL Processes & AWS Data Streams

Apache Airflow & Data Transformation

Data Governance & Compliance (GDPR, HIPAA)

Machine Learning & Statistical Analysis

Automation Testing & SDLC Management

Data Integration & Real-time Streaming

Data Modeling & Warehousing

Containerization & CI/CD Pipelines

Agile Methodologies & Version Control (Git, SVN)

DevOps Practices & Cloud Storage (AWS, Azure, GCP)

Data Visualization & Predictive Analytics

Financial & Healthcare Data Analysis

Big Data Technologies (Hadoop, Spark)

Database Management (SQL, NoSQL)

Scripting Languages (Python, Bash)

Business Intelligence Tools (Tableau, Power BI)

Risk Management & Compliance Standards (SOX)

Data Security & Privacy Measures

Collaborative Tools (Jira, Confluence)

Customer Relationship Management (CRM)

Continuous Improvement & Optimization Strategies

Problem Solving & Critical Thinking Skills

PROFILE SUMMARY

➢ **Data Engineering Expertise**: Experienced Data Engineer with 3 years of expertise in designing, implementing, and optimizing data ingestion pipelines and ETL processes. Proficient in using AWS (Glue, EMR, Redshift) and GCP (BigQuery, Dataflow) for scalable data solutions. Skilled in Apache Spark, PySpark, and Apache Airflow for large-scale data processing and workflow orchestration. Adept at real-time data streaming with Kafka, ensuring data quality, and utilizing SQL and NoSQL databases.

➢ **Collaborative Team Player**: Proven track record of collaborating across teams to enhance data availability and streamline workflows, resulting in measurable improvements in operational efficiency.

➢ **Data Governance and Compliance**: Comprehensive knowledge of data governance and compliance standards (GDPR, HIPAA), ensuring data integrity and security throughout all stages of data processing.

➢ **Statistical and Machine Learning Proficiency**: Proficient in statistical analysis and machine learning techniques, including hypothesis testing, regression, classification, A/B testing, and advanced machine learning methodologies such as ensemble learning (e.g., Random Forests, Gradient Boosting), deep learning (e.g., Convolutional Neural Networks, Recurrent Neural Networks), and unsupervised learning (e.g., clustering, dimensionality reduction).

➢ **Software Engineering Practices**: Experienced in software engineering practices, including agile methodologies (e.g., Scrum, Kanban), CI/CD pipelines using tools like Jenkins and GitLab CI, and automation testing frameworks such as Selenium and PyTest. Proficient in implementing version control with Git, practicing code reviews, and utilizing containerization technologies like Docker and Kubernetes to contribute to a robust software development lifecycle management.

➢ **Project Management Skills**: Demonstrated ability to manage complex projects from conception to deployment, ensuring adherence to timelines, budgetary constraints, and quality standards.

➢ **Communication and Leadership**: Effective communicator and team player with a passion for leveraging data-driven insights to drive business decisions and optimize organizational performance.

PROFESSIONAL EXPERIENCE

June 2021 - March 2024 | Sisense

KEY RESULT AREAS:

October 2023 - March 2024 | Data Engineer

● Spearheaded the implementation of a robust AWS-based data ingestion pipeline, optimized JSON payload transformation, and orchestrated seamless data flow from S3 to RDBMS using AWS Data streams, Firehose, Lambda functions, dbt, and Airflow.

● Facilitated the decommissioning of Gainsight data using Upsolver, employed Boto3 scripts and S3 operators in Airflow for daily data transfers, and updated final tables in the relational database.

● Streamlined storage integration, Snowpipe, tables, and models through YAML-driven macros in dbt, automated processes, and reduced manual efforts.

● Implemented AWS event notifications for real-time data streaming from the data lake to the database, ensured updates, and maintained data integrity.

SOFT SKILLS

Collaboration Skills

Time Management Skills

Interpersonal Proficiency

Leadership Competence

Decision Maker

Problem-Solving Acumen

EDUCATION

- **2022 | Master of Science - Analytics** | Harrisburg University of Science and Technology, Harrisburg PA | GPA: 3.9/4.0

- **2019 | Bachelor of Arts - Economics** | Miami University, Oxford OH | Lambda Chi Alpha Fraternity: Treasurer, Phi Chi Theta Professional Business Fraternity, Commercial Banking Club

TECHNICAL SKILLS

- **Programming Languages & Databases**: Python, Java, Scala, R, SQL (Advanced), Snowflake, Redshift, BigQuery

- **ETL Tools**: Apache Airflow, DBT, Informatica

- **Cloud Platforms**: AWS, Microsoft Azure, Google Cloud

- **Data Versioning & Management**: Git, Apache Atlas

- **Data Governance & Compliance**: GDPR, HIPAA

- **Statistics & Machine Learning**: Hypothesis Testing, A/B Testing, Bayesian Statistics, Multi-Armed Bandit, Regression, Classification, Dimensionality Reduction

PROJECT DETAILS

- **Honda Capstone**: Researched Honda's 'Fuel and Brake Pipe' supplier ecosystem, developed supplier scorecards, conducted SWOT analysis, and performed forecasting analysis. Recommendations for new partnerships were adopted by Honda to enhance their lean supply chain management.

- **Optimizing SEC Data for Stock Market Prediction**: Analyzed variables influencing stock sell or pass indices and company revenue. Utilized big data techniques and data visualization in R (GGPlot, PCA, Regression Analysis) to demonstrate findings.

August 2022 - October 2023 | Software Engineer

- Developed cloud-native version control plugins for Sisense suite, utilizing JAVA, NodeJS, Docker, and Kubernetes, which had led to a 10% monthly increase in user engagement.

- Designed and implemented a comprehensive automation testing suite using JAVA, Selenium, Cucumber, and Python, achieving 95% test coverage and reducing referential integrity errors by 85%.

- Ensured stability, interoperability, portability, security, and scalability of system architecture through verification using AWS, EKS, and Terraform. Employed Postman for streamlined API testing, ensuring robust endpoint validation and seamless integration assessment.

- Documented bugs with TestRail and promptly reported defects to developers for resolution.

June 2021 - April 2022 | Software Engineering Intern

- Assisted in developing and enhancing the automation suite for the Notebooks feature using JAVA, Selenium, and Python, which had reduced translation time by 75% and mitigated referential integrity errors by over 80%.

- Engineered Image/Document classification, SVM's, Time-Series, Sentiment Analysis, and Fourier Transform samples for the Playgrounds initiative using Python and R.

- Conducted code quality assurance activities, including identifying and resolving code conflicts using JAVA, TestRail, and Selenium.

- Collaborated with the Product team to create a community blog series titled "From BI to AI using Sisense."

November 2019 - April 2020 | 3M HIS (Contract) | Data Systems Analyst

KEY RESULT AREAS:

- Modeled and analyzed healthcare data, encompassing inpatient, outpatient, professional, and pharmacy claims.

- Collaborated with internal client teams to structure, analyze, and interpret data requests.

- Teamed up with the research team to assess analytic processes and integrate them into new and existing products.

- Developed data visualizations using graphs, charts, and tables to effectively communicate findings to a lay audience.

- Executed SQL queries utilizing joins, subqueries, and indexing to extract and manipulate data.

- Offered analytic support for clients and attended on-site client meetings as needed.

- Conducted data quality assurance activities, identifying and resolving data errors to ensure accuracy and reliability.

May 2017 - July 2017 | ICICI | Institutional Equity Sales Intern

KEY RESULT AREAS:

- Explored potential service opportunities for High Net Worth Clients, Private Equity Firms, and Family Offices in India. Conducted market research and analysis for new investments.

- Collected data and prepared portfolio performance reports for large equity investments and key companies in the Housing Finance sector. Assisted Associate and Analyst in evaluating potential stocks for investment.

- Facilitated communication with target company's finance/accounting departments and advisors to gather relevant accounting support.