**FULL NAME**

(123) 456-7890 | email@mail.com

**PERSONAL SUMMARY:**

Data Engineer with over six years of experience in data engineering, architecture, and specializing in end-to-end solution development. Proven track record of leveraging cutting-edge technologies to drive business outcomes and optimize processes across automobile, insurance, and finance industries. My expertise lies in the architecture and development of databases and implementing end-to-end ETL pipelines. I excel at architecting cloud environment to ensure optimal performance.

**PROFESSIONAL SUMMARY:**

* Designed and deployed end-to-end MLOps pipelines using GitLab CI/CD, Docker, and Kubernetes, streamlining model training, testing, and deployment processes for efficiency and reliability.
* Implemented comprehensive model monitoring systems to track performance metrics, detect data drift, and trigger timely model retraining, ensuring sustained accuracy in production.
* **Expert in automating deployment and monitoring processes** using AWS CloudFormation, CloudWatch, and Datadog, ensuring high reliability and reduced downtime.
* **Proven expertise in ETL processes**, including data ingestion, transformation, and storage optimization, ensuring efficient data flow and minimizing latency.
* **Demonstrated ability to lead end-to-end data projects**, from initial architecture and development to deployment and ongoing optimization, ensuring high availability and performance.
* **Experience in developing real-time dashboards and reports** using Tableau and Power BI, providing key insights to support strategic decision-making.
* **Strong focus on data governance and security**, implementing robust measures for data encryption, access control, and regulatory compliance.
* **Hands-on experience in predictive analytics**, collaborating with data scientists to deploy machine learning models that drive actionable business insights.
* **Strong focus on data governance and security**, implementing robust measures for data encryption, access control, and regulatory compliance.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| Languages | Python, SQL. |
| Databases | MySQL, PostgreSQL, MongoDB, Pinecone, Cassandra, AWS, Snowflake |
| Cloud Technologies | AWS, Snowflake, Databricks |
| Big Data & Data Engineering/Streaming | AWS Glue, Kinesis, Kafka, Apache Spark, MapReduce, Databricks, Snowflake, Amazon EMR. |
| Data Visualization | Power BI, Tableau, Looker, QuickSight, Domo, FiveTran, Seaborn, Matplotlib. |
| Data Science/LLM’s Packages | TensorFlow, Pytorch, NLTK, XGBoost, Pandas, NumPy, Scikit- learn, Hugging Face, OpenAI & LangChain. |
| Large Language Models/GenAI | NLP, LLM’S, fine-tuning, RAG, Llama Series, Genimi pro, Sentiment analysis, Gan’s, LORA, QLORA. |
| Deployment Techniques/MLops | Streamlit, FastAPI, Amazon SageMaker, Restful, ZenML, AutoML, MLflow, A/B testing, Docker, Kubernetes. |
| Software Version Control & Documentation | Git, Github, JIRA, AWS CodeCommit, AWS CodeDeploy. |
| Monitoring and Logging Tools | AWS CloudWatch, AWS CloudTrail, Datadog |

**WORK EXPERIENCE:**

**Sr. Data Engineer** **Jun 2022 – Present**

Honda Motor Company, Remote.

* Architected and implemented a scalable real-time data pipeline using AWS Kinesis, Lambda, and S3 to process telemetry data.
* Developed data ingestion and processing workflows leveraging Apache Kafka to stream and enrich vehicle telemetry data, ensuring real-time data availability.
* Integrated Snowflake as the central data warehouse, optimizing storage and query performance through partitioned tables for advanced analytics.
* Collaborated with data scientists to design and deploy machine learning models in Python that predict vehicle component failures, triggering proactive maintenance alerts.
* Established monitoring and log systems using AWS CloudWatch and Datadog, optimizing data flow to handle peak loads with consistent throughput and minimal latency.
* Worked with cross-functional teams, including software engineers and product managers, to define technical requirements and ensure seamless pipeline integration.
* Implemented data governance practices, ensuring data encryption, access control, and compliance with data privacy regulations across all telemetry data processes.
* Automated data pipeline workflows using AWS Glue and Lambda, reducing manual intervention and improving data freshness for real-time analytics.
* Set up a monitoring framework using Datadog to track data pipeline performance and flow, allowing for proactive issue detection and logging.
* Designed and delivered dashboards and reports in Power BI, providing insights into vehicle performance, potential failure points, and predictive maintenance effectiveness.
* Improved predictive maintenance accuracy, reducing unexpected vehicle breakdowns and enhancing customer satisfaction.
* Reduced overall maintenance costs through proactive identification and resolution of vehicle issues before they become critical.
* Achieved sub-second data processing latency, ensuring timely and accurate data for predictive analytics and decision-making.
* Managed the end-to-end data pipeline lifecycle, from initial architecture to deployment and ongoing optimization, ensuring high availability and reliability.
* Executed exploratory data analysis to uncover patterns and relationships in customer behavior, which guided feature selection and informed the development of more effective predictive models.
* Visualized prediction outcomes using Power BI, creating interactive dashboards and comprehensive reports to effectively communicate insights and predictions to key stakeholders.
* Delivered knowledge transfer sessions and detailed documentation to enable smooth handover of the solution to the operations and maintenance teams.

**Environment:** AWS Kinesis, AWS Lambda, AWS S3, Apache Kafka, Snowflake, Python, AWS Glue, AWS CloudWatch, Power BI, Machine Learning, Data Pipeline, Datadog

**AWS Data Engineer** **Jan 2020 – Jun 2022**

Berkshire Hathaway Homestate Companies (BHHC), Omaha, NE

* Architected the cloud infrastructure on AWS using AWS CloudFormation to automate the provisioning of resources like S3, EC2, RDS, Redshift, etc.
* Designed and implemented scalable data pipelines that integrated AWS services (S3, RDS, Redshift, Glue) with Snowflake to enable seamless data processing and storage.
* Architected the end-to-end data flow from AWS S3 to Snowflake, ensuring efficient data ingestion, transformation, and loading processes while minimizing latency.
* Developed custom ETL processes using AWS Glue and SQL to transform raw data into actionable insights, optimizing data for downstream analytics.
* Performed extensive data preprocessing, including data cleaning, normalization, and feature engineering, to prepare large-scale historical data for accurate analysis and predictive modeling.
* Implemented robust error handling mechanisms, including retry logic and alerting, to ensure data pipeline reliability and quick resolution of data processing failures.
* Automated the deployment and management of data pipeline components using AWS CloudFormation, reducing setup time and ensuring consistency across environments.
* Integrated AWS Lambda functions to trigger real-time data processing workflows based on specific events, improving data timeliness for business-critical operations.
* Designed the security architecture with IAM roles and policies.
* Utilized AWS KMS for encryption to protect data in transit and at rest.
* Configured AWS Redshift as a temporary staging environment for large datasets, allowing for efficient data aggregation before final transfer to Snowflake.
* Conducted performance tuning of SQL queries and Snowflake data warehouse configurations to enhance query speed and reduce processing costs.
* Set up a monitoring framework using Datadog to track data pipeline performance and flow, allowing for proactive issue detection and logging.
* Optimized data partitioning and storage strategies in Snowflake, ensuring the pipeline could efficiently handle varying data volumes without performance degradation.
* Delivered detailed documentation and knowledge transfer sessions to ensure the smooth handover of the solution to the operations team.
* Achieved near-zero downtime with the implementation of robust error handling and automated recovery mechanisms.

**Environment:** AWS Lambda, AWS Glue, Redshift, Snowflake, SQL, CloudWatch, ETL, Data Pipeline, CloudTrail, Datadog

**Data Engineer** **Mar 2018 – Nov 2019**

NerdWallet, San Francisco, CA

* Designed and optimized end-to-end ETL pipelines using AWS services (S3, Glue, Redshift, RDS), ensuring automated T-1 data extraction, transformation, and loading of data.
* Engineered and implemented an AI-powered chat/voice bot using Amazon Lex to replace the legacy IVR system, automating customer support processes and leading to a 60% reduction in manual intervention.
* Created and managed data streaming pipelines using Amazon Kinesis, enabling the company to process and analyze customer behavior data in real time, allowing the company to respond quickly to changes in customer preferences and market trends.
* Conducted extensive data analysis on customer support interactions, leveraging machine learning to extract actionable insights and inform product update strategies.
* Enhanced ETL pipelines to extract data form Salesforce email campaign data, operational data, and external sources, achieving increase in performance efficiency.
* Developed and deployed machine learning models in production utilizing AWS SageMaker and EKS to enhance data quality, achieving an improvement in overall data utilization.
* Developed live dashboards for real-time issue tracking using Power BI, enabling the software development and product teams to monitor, analyze, and address the most frequent customer complaints.
* Collaborated with marketing and customer support teams to implement targeted retention strategies based on churn predictions, directly contributing to improvement in customer satisfaction.
* Performed extensive data preprocessing, including data cleaning, normalization, and feature engineering, to prepare large-scale historical data for accurate analysis and predictive modelling.
* Improved customer satisfaction by streamlining the support process, reducing response times, and providing the development team with insights to address common issues proactively.
* **Implemented data security and privacy measures** in compliance with industry regulations, ensuring data protection and governance.
* Established more secure customer profiles using Amazon Cognito, further improving customer privacy.

**Environment:** AWS S3, AWS Glue, AWS Redshift, Amazon RDS, Amazon Lex, Amazon Kinesis, AWS SageMaker, AWS EKS, Python, Power BI, ETL, Amazon Cognito

**EDUCATION:**

* University of Texas at Dallas - Master of Science in Business Analytics; Specialization in Applied Machine Learning.