**NAME**

**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 GitHub 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.
* Architected the cloud infrastructure on AWS using AWS CloudFormation and Terraform (IAC) to automate the provisioning of resources like S3, EC2, RDS, Redshift, etc.
* 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.
* Architected the cloud infrastructure on AWS using AWS CloudFormation (Infrastructure as Code - IAC) to automate the provisioning of resources like S3, EC2, RDS, Redshift, etc.
* Hands-on experience in predictive analytics, collaborating with data scientists to deploy machine learning models that drive actionable business insights.
* Applied AWS Glue Data Catalog to organize and manage metadata, and SageMaker-Data Wrangler to process large datasets for analysis, reporting efficiently and Datadog for monitoring.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| Languages | Python, SQL. |
| Databases | MySQL, PostgreSQL, MongoDB, Pinecone, Faiss,, Cassandra, RDS, DocumentDB, Snowflake |
| Cloud Technologies | AWS, Snowflake, Databricks |
| Big Data & Data Engineering/Streaming | AWS Glue, Kinesis, Kafka, Apache Spark, MapReduce, Databricks, Snowflake, Amazon EMR, Redshift. |
| 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, NLU, LLM’S, fine-tuning, RAG, 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, GitLab, CI/CD, JIRA, AWS CodeCommit, CodeDeploy. |
| Monitoring and Logging Tools | AWS CloudWatch, AWS CloudTrail, SNS |

**Sr. Data Engineer**  **Jan 2020 – Present**

Honda Motor Company, Torrence, CA.

* Architected and implemented a scalable real-time data pipeline using AWS Kinesis, Lambda, S3, and SQS to process and onboard telemetry data, ensuring efficient real-time analytics.
* Developed data onboarding and processing workflows leveraging Managed Streaming for Apache Kafka (MSK), SNS, and PySpark to stream, enrich telemetry data, achieving real-time data availability.
* Architected cloud infrastructure on AWS using AWS CloudFormation to automate provisioning of resources like S3, EC2, RDS, Redshift, SNS, and SQS, streamlining resource management.
* Integrated Snowflake as the central data warehouse, optimizing storage and query performance through partitioned tables and SQL-based transformations for advanced analytics.
* Collaborated with data scientists to design and deploy machine learning models in Kubernetes and SageMaker, enabling scalable, containerized deployments and model predictions oncomponent failures.
* Used AWS Glue Data Catalog and SageMaker-Data Wrangler to manage metadata, process large datasets, and perform data transformations, facilitating efficient data monitoring and analysis.
* Developed and managed large-scale data pipelines using MapReduce, PySpark, and MSK for big data processing in ETL workflows, incorporating SQS to manage message queuing across processes.
* Automated deployment and management of data pipeline components using AWS CloudFormation (IaC), Kubernetes, and GitHub CI/CD ensuring environment consistency and scaling capabilities.
* Leveraged Scikit-learn, TensorFlow, and PySpark to train machine learning models on SageMaker, facilitating scalable model deployment and efficient processing of telemetry and transaction data.
* Designed and delivered Power BI dashboards for insights into vehicle performance, failure points, and predictive maintenance, supporting data-driven decisions.
* Collaborated with DevOps to establish CI/CD pipelines using GitHub, Jenkins, Kubernetes, and AWS CodePipeline, supporting efficient deployment and version control for data infrastructure.
* Deployed and maintained MLOps pipelines in production with SageMaker-Data Wrangler and Elastic Kubernetes Service (EKS), enhancing data quality and enabling model retraining.
* Managed end-to-end data pipeline lifecycle from design to deployment, ensuring reliability, scalability, and high availability for PySpark-based pipelines.
* Created custom ETL processes with AWS Glue, PySpark, SQS, and SQL to transform raw data into actionable insights, optimizing data for advanced analytics and machine learning.

**Environment:** AWS Kinesis, AWS Lambda, S3, Glue, CloudWatch, SQS, SNS, RDS, Redshift, Kubernetes, Apache Kafka, Snowflake, Python, PySpark, Power BI, GitHub, CI/CD, CloudFormation (IaC)

**Data Science Engineer**  **Aug 2016 – Jun 2018**

Southwest Airlines, Dallas, TX

* Designed and implemented scalable data pipelines using AWS Glue and PySpark, automating ETL processes for handling large volumes of structured and unstructured data, integrating them into Snowflake.
* Developed Infrastructure-as-Code (IaC) using CloudFormation to provision and manage AWS resources, ensuring consistency across environments, enhancing scalability, and reducing configuration errors.
* Built real-time data streaming applications with AWS Kinesis, MSK (Managed Streaming for Apache Kafka), and AWS SQS, processing millions of events per second for live analytics and message queuing.
* Led the migration of legacy on-premises data pipelines to the AWS Cloud using CloudFormation (IaC), achieving high availability, security, and scalability across multiple environments.
* Leveraged Kubernetes for managing distributed containers, ensuring efficient resource management and high availability, and utilized SNS and SQS for decoupled communication within architecture.
* Developed and managed data workflows with PySpark, leveraging Git for version control and GitHub CI/CD pipelines for streamlined operations across development, testing, and production environments.
* Collaborated with DevOps to implement CI/CD pipelines using GitHub, Jenkins, and AWS CodePipeline, ensuring seamless deployment and version control for infrastructure and data pipelines.
* Implemented real-time monitoring and anomaly detection with AWS CloudWatch, Amazon SNS, and Amazon SQS, creating custom dashboards and setting up alerts for proactive system health monitoring.
* Deployed containerized applications using Docker and managed orchestration through Kubernetes (EKS), facilitating efficient scaling and resource allocation for data pipelines, enhancing fault tolerance.
* Built and optimized data lakes on Amazon S3 and AWS Redshift for storing and analyzing terabytes of historical and real-time transactional data, improving data accessibility and scalability.
* Enhanced data pipeline resilience and scalability through PySpark-based batch and real-time processing workflows, optimizing performance for distributed cloud infrastructure.
* Streamlined batch and real-time data processing using PySpark to optimize distributed data pipelines in the cloud, enhancing performance and reducing latency.

**Environment**: Python, SQL, PostgreSQL, Snowflake, AWS Glue, CloudFormation, AWS Kinesis, Docker, Kubernetes (EKS), AWS CloudWatch, SNS, SQS, S3, Redshift, Git, GitHub, CI/CD, Jenkins, PySpark.

**AWS Data Engineer** **Jun 2018 - Dec 2020**

Berkshire Hathaway Homestate Companies (BHHC), Bengalore, IND

* Architected the cloud infrastructure on AWS using AWS CloudFormation (IAC) to automate the provisioning of resources like S3, EC2, RDS, Redshift, etc.
* Architected the end-to-end data flow from AWS S3 to Snowflake, ensuring efficient data onboarding, 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.
* Collaborated with DevOps to implement CI/CD pipelines using GitHub, Jenkins, and AWS CodePipeline, ensuring seamless deployment and version control for infrastructure and data pipelines.
* 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 (IAC), reducing setup time and ensuring consistency across cloud platforms.
* Integrated AWS Lambda functions to trigger real-time data processing workflows based on specific events, improving data timeliness for business-critical operations.
* Conducted performance tuning of SQL queries and Snowflake data warehouse configurations to enhance query speed and reduce processing costs.
* Optimized data partitioning and storage strategies in Snowflake, ensuring the pipeline could efficiently handle varying data volumes without performance degradation.

**Environment:** AWS Lambda, AWS Glue, Redshift, Snowflake, SQL, CloudWatch, ETL, Data Pipeline, CloudTrail, CloudFormation, Terraform (IAC)

**Data Engineer** **Aug 2014 – Jul 2016**

NerdWallet, Bengaluru, IND

* 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 reduction in manual intervention.
* 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.
* 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;   
  Graduate Certificate: Specialization in Applied Machine Learning.