RANJITH NAMALA

Fresno, CA 559-293-9778 Ranjith.chowdary.namala@gmail.com linkedin.com/in/ranjith-namala-631260260/

DATA ENGINEER

PROFESSIONAL SUMMARY

- Data Engineer with 3+ years of experience in Big Data, Cloud Engineering, and Data Warehousing across AWS and Azure platforms.
- Proficient in building robust data pipelines using Azure Data Factory, Azure Databricks, and managing storage via Azure Data Lake, SQL Database, and Data Warehouse.
- Skilled in AWS services including EC2, S3, EMR, Lambda, CloudWatch, RDS, IAM, and VPC for scalable, secure cloud infrastructure.
- Experienced in developing Spark applications using Spark SQL, Streaming, DataFrames, Datasets, and integrating with Hadoop platforms like Cloudera, EMR, and HDInsight.
- Strong programming background in Python using Pandas, NumPy, Scikit-learn, and Flask, along with PySpark, Scala, and Spark-SQL for data analysis, transformation, and machine learning.
- Expertise in HiveQL, Pig, MapReduce, and ACID-compliant Hive tables for efficient querying and big data processing.
- Proficient in data modeling (Star/Snowflake schemas), advanced SQL, Tableau dashboarding, A/B testing, and data storytelling for business insights.
- Hands-on experience in containerization and cloud deployment using Kubernetes, CloudFormation templates, and PowerShell scripting.
- Adept in Agile environments, collaborating with cross-functional teams, utilizing Git, Jira, and Remedy for version control, issue tracking, and platform support.

WORK EXPERIENCE

Data Engineer - Capitalone | ***Contract*** *San Fransisco, California* **Jan 2024 – Present**

- Designed and implemented a secure data archiving solution to transfer sensitive unstructured data including XML, JSON, PDFs, and scanned documents from Box to Amazon S3, supporting regulatory compliance and internal audit readiness.
- Developed AWS Lambda functions to extract critical metadata such as customer identifiers, file path, project codes, and document types upon ingestion into S3, enabling traceability and structured storage of financial records.
- Leveraged Amazon SQS to manage metadata workflows across systems, ensuring reliable and asynchronous processing of customer and transactional data.
- Built a resilient AWS Lambda pipeline incorporating Dead Letter Queues to ensure reliable ingestion of critical files, minimizing data loss and supporting continuity in regulatory reporting.
- Engineered a robust retry mechanism to automatically reroute failed messages, enabling compliance-grade data reliability and reducing operational risk.
- Applied intelligent retry thresholds to balance system performance with the need for exhaustive retries, ensuring uninterrupted processing of high-priority financial data.
- Configured and monitored Amazon SQS queues and DLQs to maintain visibility over archiving workflows and ensure timely processing of compliance-related files.
- Implemented end-to-end file encryption using AES 256-bit during transfers, with individualized encryption keys stored securely in compliance with banking data protection standards.
- Automated structured data migration from core banking systems to Amazon S3 using AWS DMS, optimizing data accessibility for analytics and regulatory reporting.
- Integrated AWS Glue crawlers and Athena to enable searchable, query-ready data lakes for structured data, empowering downstream teams with real-time insights for risk analysis and financial decision-making.

Data Engineer I - SNKMDR Technologies | *Chennai, India* **Jul 2021 – Dec 2022**

- Automated Outlook tasks and monitoring tools using Python, reducing manual intervention and improving operational efficiency.
- Conducted root cause analysis of production issues using Linux and SQL, leading to reduced downtime and increased system reliability.
- Implemented an S3 event-based architecture that triggered AWS Lambda functions for real-time data processing workflows.
- Developed Lambda functions to initiate Glue Crawlers for metadata extraction from S3, enabling efficient data cataloging.
- Configured Glue Crawlers to update metadata in the AWS Glue Data Catalog, streamlining data discovery and query readiness
- Established CloudWatch Event Rules to monitor Glue Crawler success and trigger downstream Lambda functions for ETL operations.
- Created Lambda functions to launch AWS Glue Jobs based on crawler success, ensuring timely transformation of ingested data
- Developed Glue Jobs to transform raw datasets into optimized formats, improving storage efficiency and analytical performance.
- Enabled workflow completion alerts through CloudWatch and Amazon SNS, incorporating email verification for secure and reliable notifications.
- Enhanced ETL pipelines by cleansing and transforming nested JSON data using Glue DataBrew, with final datasets queried in Athena and visualized in Tableau.

Data Engineer Intern | *Adani Group, chennai, India* **Feb 2021 – Jul 2021**

- Automated data validation workflows for new Redshift entries using Python, significantly improving data accuracy and quality.
- Created stored procedures and user-defined functions to optimize complex joins and aggregations, reducing average query time by over 75 percent.
- Consolidated data from multiple source systems into a Redshift data warehouse, designing a star schema and developing ETL pipelines with SQL and Python to boost query efficiency.
- Built and maintained Tableau and QuickSight dashboards to visualize KPIs and operational trends, enabling data-driven decisions and increasing revenue and stakeholder engagement.
- Conducted A/B testing on dashboard designs and visualizations, improving usability and increasing dashboard adoption by 15 percent.
- Profiled and organized new data sources using Pandas, structuring key metrics, KPIs, and dimensions to support scalable data lakes and analytics.
- Investigated and resolved data quality issues during ETL processes by debugging AWS Glue Spark jobs and analyzing system logs, enhancing pipeline reliability

ACADEMIC PROJECTS

Real estate home price prediction using Machine Learning **Aug 2023**

- Built a regression model in Python to predict house prices based on features like location, area, and number of rooms.
- Performed data cleaning, feature engineering, and exploratory analysis to identify key drivers of price variations.
- Trained and evaluated multiple machine learning algorithms using Scikit-learn, optimizing model performance through cross-validation.

Real-Time IoT Data Streaming and Analytics using AWS **Aug 2023**

- Designed and implemented a real-time data pipeline for continuous ingestion of IoT sensor data using AWS Kinesis Data Firehose.
- Enabled streaming analytics and decision-making by integrating data with Amazon S3 for scalable storage and downstream processing.
- Delivered timely and accurate insights from real-time data streams, improving monitoring and responsiveness for IoT-driven operations.

Weather Data Pipeline Using Apache Airflow and AWS S3 **Mar 2023**

- Developed an automated data pipeline using Apache Airflow to extract real-time weather data from an external API.
- Implemented data transformation workflows to clean and structure weather data before storing it securely in Amazon S3.
- Enabled scalable, reliable, and scheduled data processing, supporting analytics and timely weather-driven decision-making.

Traffic Prediction for Intelligent Transportation Systems Using Machine Learning **Aug 2023**

- Developed a machine learning model to predict traffic flow patterns using historical data and time-based features for improved transportation planning.
- Applied data preprocessing techniques and trained regression models to generate accurate, real-time traffic predictions.
- Enabled proactive traffic management by delivering timely insights to support intelligent routing and congestion reduction strategies.

Intelligent Blood Bank Management and Automation System **Jan 2023**

- Developed a system to automate donor registration, blood stock tracking, and workflow operations, enhancing efficiency in blood bank management.
- Integrated functionalities for real-time inventory monitoring, donor compatibility matching, and end-to-end traceability of blood units.
- Reduced manual effort and improved data accuracy through streamlined workflows and centralized data management.

Invoice Processing and Analytics System Using OCR and AWS **Jun 2023**

- Extracted structured data from scanned invoices using OCR techniques and automated ingestion into Amazon S3.
- Cataloged and queried invoice data using AWS Glue and Athena, enabling real-time financial analytics and reporting.

Customer Segmentation and Recommendation System Using Machine Learning **Sep 2023**

- Built a segmentation model using K-Means clustering to group customers based on behavioral and transactional patterns.
- Developed a collaborative filtering recommendation engine to personalize product suggestions and enhance user engagement.

SKILLS

Big Data & Cloud Technologies: HDFS, YARN, MapReduce, Pig, Hive, HBase, Cassandra, Oozie, Apache Spark, Impala, Kafka, StreamSets, Amazon EMR, EC2, EBS, S3, RDS, Redshift, Glue, Firehose, Lambda, IAM, Kinesis, SQS, DynamoDB, Elasticsearch, Azure HDInsight, Databricks, Data Lake, Blob Storage, Data Factory, SQL Database, Data Warehouse, Cosmos DB, Azure DevOps, Active Directory
Programming Languages & Scripting: Python, Scala, Java (J2EE), SQL, HiveQL
Databases: Snowflake, MySQL, PostgreSQL, MS SQL Server, Oracle, Teradata, MS Access, DB2
ETL & Reporting Tools: Informatica, Talend, SSIS, SSRS, SSAS, Tableau, Power BI, ER Studio
Development & DevOps Tools: Eclipse, IntelliJ, NetBeans, Jupyter Notebook, Docker, Kubernetes, Jenkins, Jira, Microsoft Office Suite
Methodologies: Agile, Scrum, Waterfall
Other Skills: Machine Learning, Natural Language Processing (NLP), Spring Boot, Linux

EDUCATION

California State University, Fresno **Dec 2024**
Master of Science - Computer Engineering *Fresno, California, USA*

Veltech University **May 2022**
Bachelor of Technology - Electronics and Communications Engineering *Hyderabad, India*

ACADEMIC EXPERIENCE

Teaching Assistant | *California State University, Fresno* **Aug 2023 – Dec 2023**

- Assisted students in Introduction to Machine Learning with implementation of core ML algorithms and hands-on projects involving data preprocessing, model training, and evaluation using Python and Jupyter Notebooks.
- Guided students on C++ and Python scripting for system validation, hardware debugging, and efficient memory management in applied computing environments.