BRAJESH KUMAR JHA

+91 7484837464 | brajeshjha2001@gmail.com| linkedin.com/in/brajeshkrjha

PROFESSIONAL SUMMARY

- Cloud Data Engineer with 2 years of experience in data transformation, cloud migration, and data lake implementation.
- Certified AWS Cloud Practitioner with a strong foundation in cloud architecture and data transformation best practices.
- Solid understanding of the U.S. insurance domain, including personal and commercial auto, and workers' compensation lines of business (LOBs).
- Skilled in leveraging AWS services (S3, Lambda, Glue, Athena, SQS, EventBridge, Step Functions, CloudWatch), MarkLogic, and Terraform to design scalable data solutions.
- Experienced in migrating legacy systems to the cloud, automating data workflows, and optimizing performance in high-volume environments.
- Proficient in PySpark for ETL processes, handling complex XML data transformations, and implementing Apache Iceberg for data lake solutions.
- Experienced in the software development life cycle (SDLC), Test-Driven Development (TDD), and Agile methodologies.

EDUCATION

- **Bengal Institute of Technology, Kolkata (WB)**
  Bachelor of Technology – Information Technology | CGPA – 9.01 *August 2019 - June 2023*

TECHNICAL SKILLS & CERTIFICATIONS

- Certifications: AWS Certified Cloud Practitioner
- Programming & Scripting: Python, PySpark, XQuery, SQL, HCL.
- Cloud Technologies: AWS (S3, Lambda, Glue, Athena, SQS, DataSync, EventBridge, Step Functions, SNS, Lake Formation, KMS, IAM, CloudWatch, CloudTrail).
- Tools & Frameworks: Terraform, AWS SDK, CI/CD pipelines (GitLab CI/CD).
- Big Data and ETL: Data Lakes, Apache Iceberg, AWS Glue, Event-Driven Architectures, Data Transformation.
- Database Technologies: MarkLogic, MySQL, DynamoDB.
- Data Warehousing & Analytics: Amazon Athena, Amazon Redshift.
- Domain Knowledge: US Auto Insurance Property and Casualty
- Data Formats: XML, Parquet, JSON, CSV.

WORK EXPERIENCE

Cloud Data Engineer
ValueMomentum Software Services Pvt Ltd *Jan 2025 - Present*

- Contributed to migrating the project from MarkLogic to AWS by designing and building a scalable data lake solution to store and process large volumes of XML data, converting the data into Parquet files for optimized querying and analytics.
- Implemented automated data pipelines using AWS services (S3, SQS, Lambda, EventBridge, Step Functions, DataSync, CloudTrail, Glue, Athena, IAM, CloudWatch) to transform XML data into Parquet format, ensuring efficient and reliable data processing.
- Utilized Terraform to build and automate the end-to-end infrastructure, enabling consistent, scalable, and secure deployments across AWS services while reducing manual intervention.
- Integrated Apache Iceberg to support time travel, schema evolution, and ACID transactions within the data lake.

Software Engineer - Trainee
ValueMomentum Software Services Pvt Ltd *Aug 2023 – Dec 2024*

- Contributed to the development of an Enterprise Data Hub in Amazon Redshift by using MarkLogic to harmonize segmented XML files into a unified policy XML format before loading the transformed data into Redshift.
- Leveraged MarkLogic CoRB to convert XML data into CSV files, creating fact and dimension tables in data marts for seamless querying in Amazon Redshift.
- Ensured data consistency between MarkLogic XMLs and Redshift, achieving over 95% accuracy during the transformation process and optimizing data for high-performance analysis and reporting.
- Automated batch jobs using shell scripts and Gradle tasks to upload CSV files to Amazon S3, streamlining integration with Redshift and minimizing manual intervention.

PROJECTS

Commercial Prefill Data Lake

- Designed and implemented a scalable data lake architecture on AWS to store and process commercial insurance prefill data.
- Developed ETL pipelines using AWS Glue and PySpark to transform raw data into a structured format for analytics.
- Implemented data quality checks and monitoring using AWS CloudWatch and Lambda functions.

MarkLogic Data Hub Framework

- Designed and developed data harmonization flows to standardize data from multiple source systems.
- Implemented data quality rules and validation processes to ensure data integrity.
- Developed APIs to expose the harmonized data to downstream applications and services.