

Mr. Bhushan Bamane

AWS Data Engineer

bhushanbbamane@gmail.com

+91 -7378382661

PROFESSIONAL SUMMARY

- Having **4+ year** of experience as a Cloud Data Engineer in **Amazon Web Services (AWS)**, **Spark**, **SQL**, **Python** and **Hive**.
- Having experience in various AWS Services like S3 bucket, Redshift, EMR, Glue, Athena, Lambda, EC2, RDS, Cloud Watch, SNS, IAM, KMS.
- Worked on creation of data pipelines using **Apache Airflow** and **AWS Lambda**.
- Skilled in **Amazon S3** for storing and retrieving data efficiently.
- Proficient with **EC2** instances for seamless scaling and support for various instance types.
- Experienced in creating Lambda functions for event-triggered **ETL** processes and data ingestion.
- Knowledgeable in query engines and schema catalog tools like Athena and Glue for analytical queries and transformations.
- Hands-on experience with **PySpark** scripts in AWS **Glue** for ETL leveraging distributed computing capabilities.
- Expertise in building table definitions and working with different file formats in S3 for efficient data extraction and transformation.
- Optimizing Spark queries in AWS **EMR** for improved performance using partitioning, caching, and tuning configurations.
- Configuring and utilizing AWS **CloudWatch** for resource monitoring and application metrics to enable proactive scaling and automated responses.
- Expertise in leveraging **Amazon Redshift** capabilities to facilitate **OLAP**-based analysis, enabling efficient processing of complex queries and multidimensional data exploration for insightful decision-making.
- Good knowledge of **RDS Oracle** for **OLTP** processes data with a focus on high scalability and low latency.
- Developed data processing workflows using Apache Spark, achieving a improvement in data processing speed and Proficient in writing PySpark scripts to apply client-specific transformations on data.
- Experienced in optimizing Spark queries within AWS EMR to enhance performance.
- Strong ability to analyze data using PySpark, enabling valuable insights and data-driven decisions.
- Skilled in ETL operations on data frames using PySpark, ensuring efficient data processing.
- Extensive experience with Hive, including partitioning, bucketing, and various join types on Hive tables.
- Familiarity with different file formats (CSV, Parquet, Avro, ORC, JSON) in Hadoop ecosystem.
- In-depth understanding of Spark architecture, RDD, Master and Worker nodes and Hive architecture.
- Proficient in SQL for data transformation and building optimal analytic queries.
- Expertise in using Airflow for orchestrating Spark jobs and developing data pipelines for various AWS services like S3, Glue, EMR, Redshift.

EXPERIENCE DETAILS

Working as a AWS Data Engineer at **Ecojoy Tech Pvt Ltd, Pune** from **Sep - 2019 to till date**

TECHNICAL SKILLS

- **Amazon Cloud Services:** S3, EC2, Glue, EMR, Athena, Redshift, Lambda, RDS, Airflow, CloudWatch, SNS, IAM
- **Version Control Software :** Git, GitHub
- **Scheduling Tools :** Apache Airflow
- **Programming Language :** Python
- **Hadoop Ecosystem :** Spark, Hive
- **Databases :** Amazon RDS, Oracle, MySQL
- **Data Warehouse :** Redshift
- **Tool and Software :** Pycharm, VSCode, Jupyter, Thonny
- **OS :** Windows, Linux

PROJECT DETAILS

Client	Booktopia
Description	Booktopia Group Ltd is an Australian online bookstore that was established in 2004. It has a yearly turnover of \$165 million and has been featured eight times in the AFR/BRW's Fast 100, a remarkable achievement that no other company has accomplished, from 2009 to 2017.
Role	AWS Data Engineer

ROLES AND RESPONSIBILITIES

- Design, implement, and maintain end-to-end data pipelines for data extraction, transformation, and loading (ETL) using AWS services like AWS S3, Glue, Redshift.
- Implement data quality checks (Null check, duplicate check, filling missing Values), data validation routines, data auditing and error handling mechanisms in the ETL processes.
- Implemented SCD Type 2 functionality.
- Automated processes through scripts and scheduled developed jobs on daily automated schedules using Apache Airflow.
- Worked with PARQUET, CSV file formats for data processing.
- Contributed to the AWS architecture implementation for a complete cloud-based big data solution involving Redshift, EMR, Glue, Athena, and Airflow.
- Optimize Spark job configurations and DPU settings on Glue for optimal data processing performance.
- Responsible for analysis, design, and testing phases, as well as documenting technical specifications.
- Performed pre-processing and data cleaning using Spark and AWS services like Athena and Glue.
- Utilize AWS Glue Data Catalog to manage metadata, ensuring proper data lineage, and enabling efficient data discovery.
- Implemented monitoring and alerting using CloudWatch and SNS to enable proactive detection of performance issues and resource utilization.
- Managed debugging and troubleshooting problems throughout the data processing workflow.
- Viewed logging information and initiated new tasks or resumed tasks that have failed.
- Participated in daily stand-up meetings (Scrum) and requirement review meetings to analyze requirements for each story card in a sprint.
- Responsible for pull, push, and committing the code using GitHub.

Client	Vertex Pharmaceuticals
Description	Vertex Pharmaceuticals is a biopharmaceutical company founded in 1989 and headquartered in Boston, Massachusetts. It was one of the pioneers in the biotech industry to adopt a rational drug design approach, instead of relying on combinatorial chemistry.
Role	AWS Data Engineer

ROLES AND RESPONSIBILITIES

- Developing and deploying data ingestion pipelines, utilizing AWS Glue for seamless ETL processes to transform and prepare data.
- Provisioning and optimizing Amazon Redshift clusters, designing efficient data models to support high-performance querying.
- Leveraging PySpark and SQL for robust data processing, constructing data pipelines that move data from AWS S3 to Redshift for advanced analytics.
- Implementing AWS Lambda functions for event-driven processes, triggering data processing workflows based on events like new file uploads to AWS S3.
- Implementing security measures to protect sensitive pharmaceutical data stored in S3 and Redshift, ensuring compliance with industry regulations and data privacy policies.
- Implemented SCD Type 2 functionality.
- Responsible for debugging and troubleshooting during data processing.
- Responsible for pull, push and committing the code using github.

Client	Big W
Description	Big W is a department store chain that offers a variety of products, including clothing, health and beauty items, garden supplies, kitchenware, toys, pet supplies, office supplies, books, DVDs, CDs, televisions, gaming consoles, video games, furniture items, snack food, and small electrical household appliances.
Role	Data Engineer

ROLES AND RESPONSIBILITIES

- Designing, developing, and upkeeping data pipelines utilizing Python for the extraction, transformation, and loading of data from diverse origins like databases, APIs, and files, subsequently transferring it into Amazon S3.
- Updated existing pyspark code and tried to make them more generic across the platform.
- Develop SQL Queries for fetching the data from database.
- Sorting the extracted data in Oracle database for further querying and analysis.
- Responsible for debugging and troubleshooting during data processing.
- Coordinating with cross-functional teams such as data engineering and data analytics to grasp the project's workflow and meet project prerequisites.
- Involved in daily Scum meetings.

ACADEMICS QUALIFICATION

Bachelors of Engineering

B.E. completed from Shivaji University Kolhapur Year of 2019.