Azeem Khan
Senior Data Engineer
Australia Work Permit for 2 years
+61-432928191   azeemnama@gmail.com

Results-driven Data Engineer with 6+ years of experience in PL/SQL and a strong focus on ETL processes, data extraction, transformation, and loading. Proficient in AWS and Azure technologies, including Glue, Lambda, Redshift, Azure Databricks, and Azure Data Factory, with a solid foundation in Python programming. Skilled in designing and optimizing data pipelines, ensuring data integrity and facilitating data-driven decision-making.

Key Contributions:

- **ETL Processes & Cloud Migration:** Spearheaded Glue script development with PySpark and SQL, managing the migration of critical daily jobs, extracting data from a centralized location in BASF Germany, and loading it into the master data database for the product "Field Manager."
- **Data Integration & Pipeline Management:** Implemented various joins and utilized in-memory tables in PySpark with a custom framework for maintainable code. Translated graphical representations of pipelines in Talend software to code representations in AWS Glue and Azure Data Factory.
- **Data Cleaning & Workflow Automation:** Utilized EC2 instances to run a custom .jar file for data cleaning before loading data into the main database. Automated data workflows using AWS Glue jobs and Azure Logic Apps, enhancing data processing efficiency.
- **Analytics & Visualization:** Developed and maintained standardized dashboards for customers using Power BI and Amazon QuickSight, providing actionable insights through interactive data visualizations.
- **Data Quality & Performance Optimization:** Conducted data validation and quality checks, identifying and rectifying data anomalies to maintain data integrity. Optimized SQL queries and maintained database performance, achieving a 20% reduction in query execution time.
- **Custom Solutions & Documentation:** Led the design and development of custom Oracle Applications solutions, leveraging Oracle Forms, Reports, and PL/SQL. Created and maintained comprehensive technical documentation, facilitating knowledge transfer and adherence to best practices.

Proven track record of successfully delivering high-quality, scalable data solutions to meet business objectives. Experienced in collaborating with cross-functional teams, including data scientists, analysts, and stakeholders, to address data-related challenges and drive innovation. Passionate about continuous learning and eager to contribute technical expertise to further organizational growth and success.

EDUCATION DETAILS:

- **Master of Technology,**   **October 2021 - August 2023**
  Information Technology
  University of Hyderabad, Hyderabad
  CGPA : 8.58

- **Bachelor of Technology,**   **July 2015 - May 2019**
  Computer Science and Engineering
  Integral University, Lucknow
  CGPA : 8.97

PROFESSIONAL WORK EXPERIENCE:

EPAM (Astrazeneca Client)   **September 2024 to Present**
Senior Software Engineer - Data
Responsibilities: -
- Creating Data Ingestion Pipelines on Apache Airflow and Fivetran.
- Managing a team of 5 people responsible for a total of 32 Ingestion Pipelines for Astrazeneca.
- Manage the data in data lake, so that clean and sorted data is available for end customers.
- Many R&D data needs to be cleansed and pushed to the lake house and further transformations applied using pyspark and pandas as and when needed.
- Recreated an end to end pipeline including Starburst, Collibra to catalog technical metadata from S3 into Collibra.
- Direct interaction with clients and understand the functional and non functional requirements.
- Unit level testing and debugging of any errors in the pipelines due to internal or external factors.
- Implement cataloguing of business as well as technical metadata into Collibra Environment.
- Restructured the S3 buckets and created glue tables for the data in the buckets for link with Starburst tool, creating a data product for all the data sources each, which helps a lot of other teams in the company to query the data without manually going back to S3 files being in SAS/ CSV/ JSON/ XML/ AVRO/ PARQUET etc formats.

Environment: JIRA, AWS Glue, AWS Lambda, SQS, SQL, Python, Pandas, Pyspark, Airflow, Fivetran, Talend, Data Ingestion, Data Cleansing, Datalake, Data Pipelines.

BASF, Germany (Noida Office) (Client to Cosmic IT, India)   **October 2023 to August 2024**
Senior Data Engineer
Responsibilities: -
- Worked on Glue scripts development with Pyspark and SQL.
- Worked on writing scripts that parse json received from an endpoint URL and transform and load it into the BI data warehouse in AWS.
- Migrated major daily jobs, extracting data from a centralized location in BASF Germany and loading it into the master data database for the product "Field Manager." This product contains details of all fields, crops, and recommended fertilizers.
- Implemented various joins, utilizing in-memory tables in Pyspark with a custom framework for maintainable code.
- Translated the graphical representation of the pipeline in Talend software to code representation in AWS Glue.
- Utilized an EC2 instance to run a custom .jar file for data cleaning before loading data into the main database.

Environment: AWS Glue, AWS Redshift, Amazon EMR, Amazon S3, Talend, EC2 instances, Pyspark, SQL, JIRA.

Navikenz, Hyderabad, India   **September 2022 to October 2023**

Data Engineer
Responsibilities: -
- Involved in various phases of the Software Development Life Cycle (SDLC) of the application.
- Collaborated on ETL tasks, maintaining data integrity and pipeline stability.
- Prepared written summaries to accompany results and maintain documentation.
- Contributed to an accelerator project that helped in easy data migration from on-premise to various cloud platforms, including GCP, AWS, Snowflake, and Microsoft Azure Databricks.
- Developed, implemented, and maintained data analytics protocols, standards, and documentation.
- Designed a Python-based tool to convert SQL queries from Oracle/SQL Server to Redshift-based SQL queries and designed data models for complex analysis needs.

- Created Azure Data Factory pipelines for orchestrating data movement and transformations, integrating with Azure SQL Database and Azure Blob Storage.
- Implemented data processing and analysis workflows using Azure Databricks and Azure Synapse Analytics for scalable and efficient data handling.
- Successfully created standardized versions of Spend Analytics and similar dashboards for customers in Power BI as well as Amazon QuickSight.

Environment: Azure Data Factory, Azure Databricks, Azure Synapse Analytics, Azure SQL Database, Azure Blob Storage, AWS Glue, AWS Redshift, Amazon EMR, Amazon S3, Microsoft Power BI, Amazon QuickSight, SQL Developer, PyCharm, Python Notebook, EC2 instances.

Wipro Technologies, Hyderabad, India   **July 2019 to August 2022**
Associate PL SQL Developer - Data Engineer
Responsibilities: -

- Developed and managed staging and production tables of the internationally centralized relational database using queries, subqueries, nested queries, unions, and the respective aliases.
- Created and automated monitoring and reporting jobs for Quality Assurance with production support to enhance the quality of service in Oracle Applications.
- Reported and reviewed error logs and job failures, facilitated monthly audits, and maintained SLA requirements. Developed complex PL/SQL Queries with cursors, procedures, functions, sequences, and triggers. Implemented data migration strategies to move on-premise data to AWS cloud infrastructure, including leveraging Microsoft Azure Databricks and Azure SQL Database for ETL processes.
- Distributed tasks and enhanced team collaboration under team management.
- Analyzed business requirements and transformed them into solutions for AWS and Azure cloud-based applications used in the industry.
- Developed ETL processes utilizing AWS Glue and Azure Data Factory to transform and load data from various sources into Redshift and Azure Synapse Analytics.
- Automated data workflows using AWS Glue jobs, Azure Logic Apps, and Azure Functions, enhancing data processing efficiency.

Environment: Oracle Applications, AWS Glue, AWS Lambda, AWS Redshift, AWS CloudWatch metrics, Azure Data Factory, Azure Logic Apps, Azure Functions, Azure Databricks, Azure SQL Database, Azure Synapse Analytics, Oracle Cloud API, Coupa Procurement Tool, Trintech Cadency, Jira, Uber's custom software: Ucash.

CERTIFICATION DETAILS:

- **AWS Data Analyst Specialty**
  May 2023
- **Analyzing and Visualizing Data with Microsoft Power BI**
  July 2023
- **Oracle Cloud Infrastructure Foundations**
  November 2020
- **Oracle Financials Cloud 2018 Solution Engineer Specialist**
  November 2020
- **Microsoft Azure Fundamentals AZ-900**

August 2020

REFEREE DETAILS:

- **Vipul Panchal**
  Software Development Engineer (Amazon)
  Ex-Colleague (Wipro Limited)
  +91-8922827288
  vipulpanchal372@gmail.com

- **Dr Rukma Rekha**
  Professor
  University of Hyderabad
  +91-9885333934
  rukmarekha@uohyd.ac.in