**Rakesh Boda**

[LinkedIn](http://www.linkedin.com/in/boda-rakesh-0922011a4) | [GitHub](https://github.com/Rakesh2273) | [Portfolio](https://rakesh2273.github.io/Portfolio/) | [Medium](https://medium.com/@bodarakesh001)

[bodarakesh001@gmail.com](mailto:bodarakesh001@gmail.com)

+1 (203)-936-9025

**Professional Summary**

* Experienced **Data Engineer** and **Data Analyst** with a strong background in designing and developing scalable **ETL pipelines**, **data architectures**, and **analytics solutions**.
* Expertise in **Python, SQL,** and **Apache Spark (PySpark)** for big data processing, transformation, and analytics.
* Skilled in **data modeling, database design, and performance optimization** in **PostgreSQL, Redshift, and SQL Server**.
* Hands-on experience with **cloud platforms**, including **AWS services (S3, Redshift, Athena, Glue)** for data storage, processing, and analytics.
* Proficient in **building and optimizing ETL workflows** using **PySpark, Airflow, and SQL Alchemy** for real-time and batch data processing.
* Built and managed **data lakes**, integrating **structured and semi-structured data** for **business intelligence and reporting**.
* Designed and implemented **RESTful APIs** to facilitate **data integration across multiple platforms**.
* Created **interactive Power BI and Tableau dashboards** for **real-time decision-making in risk management and customer analytics**.
* Implemented **CI/CD pipelines** using **Bamboo, Bitbucket, Azure DevOps, and GitHub** to automate deployment and integration.
* Strong problem-solving abilities in **debugging software, optimizing queries, and improving data consistency across multiple sources**.
* Experience in developing and optimizing **streaming data pipelines** using **Kafka and Spark Streaming**.
* Worked on **cloud migration projects**, transitioning **on-premises systems** to **AWS and Snowflake**.
* Developed and optimized **batch job scheduling** and **workflow automation** using **Apache Airflow**.
* Built **real-time anomaly detection systems** for **fraud analytics** using **machine learning models and statistical techniques**.
* Implemented **role-based access control (RBAC)** and **security best practices** in **data governance and compliance**.
* Expertise in deploying and managing **containerized applications** using **Docker and Kubernetes**.
* Collaborated with **cross-functional teams**, including leads and managers, to develop **end-to-end data solutions**.
* Strong knowledge of **Agile methodologies**, participating in **daily stand-ups, sprint planning, and retrospectives**.

**Technical Skills**

|  |  |
| --- | --- |
| Programming Languages: | Python (Core Python, OOP, Data Structures, NumPy, Pandas, SciPy) |
| Frameworks & Libraries: | PySpark, Flask, FastAPI, Django, Pandas, NumPy, SciPy, Pytest, Unittest |
| Data Engineering: | SQL, NoSQL (PostgreSQL, MongoDB), ETL Development, Data Pipelines, AWS services (S3, Redshift, Athena, Glue) |
| Automation & Scripting: | Unix Commands, Shell Scripting, Regular Expressions |
| Big Data & Streaming: | Apache Kafka, AWS Kinesis, Dremio, Attunity Replicate |
| Performance Optimization: | Multithreading, Multiprocessing, Asynchronous programming, caching |
| Version Control & CI/CD: | REST APIs, Data Serialization, AWS Lambda |
| BI & Analytics: | Power BI, Tableau, Data Warehousing |
| Testing & Debugging: | Unit Testing (Unittest, Pytest), Exception Handling, Logging, Regression Testing |
| Cloud & Virtualization: | Docker, Kubernetes, Jenkins, GitLab |

**Professional Experience**

**Client: Citi Bank April 2024 – Present**

**Role: Data Engineer**

**Responsibilities:**

* Developed high-performance data pipelines to handle complex **financial transactions**, ensuring **minimal processing latency**.
* Designed and implemented **data lake architectures** to consolidate structured and semi-structured data from multiple sources.
* Collaborated with **risk management** teams to build predictive models for **fraud detection** and **risk assessment**.
* Integrated advanced machine learning techniques into **ETL workflows** to automate anomaly detection.
* Engineered a **data auditing framework** to ensure compliance with **banking regulations** and industry best practices.
* Worked on feature engineering and **transformation pipelines** to support **AI-driven** financial forecasting models.
* Deployed serverless data processing solutions using **AWS Lambda**, reducing infrastructure costs.
* Designed and implemented a **metadata-driven** ETL framework to enhance **data lineage** and **traceability**.
* Conducted A/B testing and impact analysis for new data pipeline enhancements.
* Automated performance monitoring for data jobs, proactively identifying bottlenecks and failures.
* Partnered with **cybersecurity** teams to implement encryption and **secure** data transfer mechanisms.
* Integrated Citi’s customer analytics data with third-party financial intelligence platforms for deeper insights.
* Conducted root cause analysis of **data discrepancies** and established **automated reconciliation** processes.
* Led the documentation of data architecture, ensuring clarity for compliance and internal audits.
* Provided training sessions for analysts and engineers on **data pipeline** best practices and **governance standards**.

**Client: Verizon January 2023 – March 2024**

**Role: Data Engineer**

**Responsibilities:**

* Developed and maintained scalable **ETL pipelines** for processing large-scale telecommunications data using **Apache Spark and SQL**.
* Integrated diverse data sources into **Verizon’s data lake**, ensuring efficient data ingestion and transformation.
* Built automated workflows **using Apache Airflow**, reducing **data processing** time by 30%.
* Designed and optimized data models in **PostgreSQL** and **Snowflake** for improved analytics and reporting.
* Created **RESTful APIs** for seamless integration with various business applications and reporting tools.
* Developed real-time data validation scripts in Python to improve data integrity and accuracy.
* Engineered batch and streaming data pipelines using **Kafka and Spark Streaming** for real-time analytics.
* Deployed cloud-based data solutions using **AWS services**, including **S3, Glue, and Redshift**, enhancing data accessibility and scalability.
* Built interactive dashboards **in Power BI and Tableau**, providing actionable insights into network performance and customer behavior.
* Collaborated with cross-functional teams to define data governance policies and best practices.
* Implemented **CI/CD** pipelines using **Azure DevOps** to streamline deployment and monitoring of data pipelines.
* Conducted performance tuning and troubleshooting to improve query efficiency and system reliability.

**Client: Hexaware Technologies October 2021 - December 2022**

**Role: Data Analyst**

**Responsibilities:**

* Conducted **data profiling, cleansing, and transformation** to improve data quality and ensure accuracy in business intelligence reports.
* Developed and maintained **ETL workflows** to streamline data ingestion from multiple sources, optimizing performance and reducing processing time.
* Created **interactive dashboards in Power BI and Tableau**, providing data-driven insights for operational efficiency and business strategy.
* Performed **exploratory data analysis (EDA) using Python and SQL** to identify trends, patterns, and anomalies in large datasets.
* Designed **SQL queries and stored procedures** to extract, transform, and load data into structured databases for advanced analytics.
* Worked closely with stakeholders to understand business requirements, delivering **customized data solutions** that enhanced decision-making.
* Automated **report generation and data validation scripts**, reducing manual effort and improving efficiency.
* Assisted in **predictive modeling and statistical analysis**, supporting risk assessment and performance forecasting.
* Collaborated with cross-functional teams, including **data engineers, business analysts, and product managers**, to develop data-driven solutions.
* Implemented **data governance best practices**, ensuring compliance with industry regulations and company policies.

**Client: LTI Mindtree January 2020 – September 2021  
  
Role:** **Software Developer**

**Responsibilities:**

* Led the development and optimization of web applications using **Python and React.js**, focusing on scalability and performance.
* Designed and implemented **RESTful APIs** for integration with various services, ensuring seamless communication between **frontend and backend**.
* Developed automated testing scripts and frameworks using **Python and Selenium,** improving test coverage and reducing manual testing efforts.
* Collaborated with **DevOps** teams to implement **CI/CD** pipelines using **Jenkins and GitLab** for automated code deployment.
* Played a key role in the design and implementation of **PostgreSQL and MySQL** databases, ensuring data integrity and optimizing query performance.
* Mentored junior developers, providing guidance on coding best practices and troubleshooting complex issues.
* Worked on performance **optimization** by analyzing and **enhancing application load times** and response times.
* Actively participated in **Agile Scrum** ceremonies, including daily stand-ups, **sprint planning**, and retrospectives, ensuring timely project delivery.
* Contributed to the overall system architecture by identifying areas for improvement and providing recommendations for software enhancements.

**Client: LTI Mindtree May 2019 - December 2019**

**Role:** **Software Developer Intern**

**Responsibilities:**

* Assisted in the development and maintenance of **web applications**, utilizing **Python and JavaScript.**
* Contributed to **backend API** development using **Flask** and integrated them with frontend applications built with **React.js**.
* Worked with the team to troubleshoot and resolve issues related to **application performance and security.**
* Assisted in developing automated tests using tools like Selenium to ensure application **functionality and stability**.
* Participated in **database design discussions**, focusing on optimizing queries and indexing for faster data retrieval.
* Supported the integration of **third-party services and APIs** into the application.
* Gained hands-on experience with version control tools like Git and participated in code reviews to ensure adherence to development standards.
* Collaborated with **cross-functional teams** to deliver software solutions in an **Agile environment.**

**Education**

**Master of Science, Business Analyst** | University of New Haven, USA