

# Akanksha Ashokrao Bable

845-853-4508 | [akankshabable1@gmail.com](mailto:akankshabable1@gmail.com) | [LinkedIn](#) | New Paltz, NY, USA

## SKILLS

---

- Relational Databases: MySQL | Microsoft SQL Server | PostgreSQL | Amazon Aurora | MS Access
- Non-Relational Databases: Object Storage | Document Store | Key-Value Store | Graph Databases | Column-Family Databases
- Languages and Frameworks: Python | Scala | Java | SAS | MATLAB | R | Stata | Apache Spark | Hadoop | Kafka | Flink | Hive
- AWS Technologies: Redshift | S3 | AWS Glue | EMR | Kinesis | FireHose | Lambda | IAM roles and permissions
- Data Visualization Software: QuickSight | Tableau | Power BI | MS Excel | PowerQuery | Macros
- Development and DevOps Tools: GitLab | Bitbucket | Docker | Jenkins | Kubernetes | Linux | Bash | Git Version Control
- Concepts: Data Modeling | Data Warehouse | Data Lake | Data Security | Data Governance | Data Quality
- Soft Skills: Problem-solving | Attention to Detail | Analytical Skills | Written and Verbal Communication Skills | Data Storytelling

## EXPERIENCE

---

### ASSOCIATE DATA ENGINEER, ACCENTURE, INDIA

December 2020 - August 2022

- Implemented and maintained data architecture using Apache Spark and AWS big data technologies to process up to 60 TB of data weekly on distributed computing systems, to support machine learning model training and meet business objectives.
- Executed ETL migration with AWS Glue, transferring 100 TB of data into Snowflake; streamlined in-depth data analysis and collaborated with a cross-functional team to optimize transformation processes, enhancing performance, scalability, and security.
- Conducted thorough code reviews for data processing and extraction scripts, identifying and resolving inefficiencies, ensuring adherence to best practices, and maintaining high standards of code quality and reliability.
- Performed in-depth analysis of 400+ datasets using Spark SQL and statistical techniques, collaborating with stakeholders to validate their integration into ETL workflows, aligning with business objectives and enhancing data-driven decision-making.
- Achieved improvements in AI/ML model performance by creating 16 new metrics using PySpark, enhancing cost analysis, operational efficiency, lookup frequency, and unit metrics, leading to an increase in prediction accuracy.

### DATA ENGINEER TRAINEE, ACCENTURE, INDIA

June 2020 - November 2020

- Optimized ETL pipeline efficiency by creating automated reporting and alert systems with 140 PostgreSQL scripts, ensuring timely notifications of data arrival from multiple sources in the data lake, and triggering downstream processes.
- Enhanced system reliability and performance by optimizing data pipelines with data compression, caching, and rigorous validation checks, implementing unit and integration testing, thereby reducing data processing time by 180 minutes.
- Designed and implemented data analytics dashboards using Tableau and QuickSight to visualize data insights stored in Redshift, reducing the analysis time of financial and operational metrics by over 10 times.

### DATA ANALYST INTERN, DATA SOLUTION, INDIA

October 2018 - April 2019

- Reduced query load times and improved query performance by 30% by applying data modeling and optimization techniques to optimize PostgreSQL database performance, utilizing targeted indexing and partitioning.
- Enhanced data quality and accuracy by implementing best practices in exploratory data analysis, cleaning, integrity checks, validation, privacy and documentation, thereby supporting critical business processes.
- Monitored ETL processes in real-time, debugging and troubleshooting in production and development, ensuring data integrity with robust error handling and automated recovery, and conducting ad-hoc analysis for seamless operations.

## PROJECTS

---

### Real-Time E-Commerce Analytics Platform

- Developed a real-time data pipeline using Apache Kafka, AWS Kinesis, and Spark, processing 10,000 simulated transactions.
- Optimized data storage with Amazon Redshift and DynamoDB, enabling efficient retrieval of 5,000 queries per hour.
- Designed a real-time sales dashboard in Amazon QuickSight, integrating automated alerts via AWS Lambda and SNS, facilitating proactive analysis of key metrics for research purposes.

### AI-Powered Sales Insights

- Engineered a data pipeline integrating 100,000+ product reviews, 50,000+ social media posts, and 5 years of sales records using AWS S3 and Glue, enabling comprehensive sentiment analysis.
- Created an interactive analytics dashboard with Amazon QuickSight, visualizing customer sentiment and predicting sales trends.

## EDUCATION

---

### Master of Science in Computer Science, State University of New York at New Paltz (GPA: 3.66/4)

August 2022 - May 2024

Coursework: Artificial Intelligence, Big Data, Cloud Computing, Data Analytics, Data Science, Machine Learning, DevOps

### Bachelor of Engineering in Computer Engineering, Savitribai Phule Pune University (GPA: 3.56/4)

August 2016 - May 2020

Coursework: Data Structures and Algorithms, Operating Systems, Software Testing and Quality Assurance, Business Intelligence