HARSH BHOLA

32-A Laxmi Nagar, Ambala Cantt, Haryana

📞 74048-88499 ✉ hb755037@gmail.com in https://shorturl.at/jLy2i ○ https://github.com/harsh0895

Education

Sanatan Dharma College   **Sep. 2021 – May 2024**
Bachelor of Computer Applications (BCA)   *A/cantt, Haryana*

S.D. Senior Secondary School   **April 2020 – March 2021**
I scored 91% marks in 12th class in Non.Medical   *A/cantt, Haryana*

Relevant Coursework

- Data Structures
- Algorithms Analysis
- Internet Technology
- Computer Architecture
- Software Methodology
- Database Management
- Systems Programming
- System Design

Experience

Complere Infosystem Pvt. Ltd.   **July 2024 – Present**
Data Engineer - I   *Ambala Cantt, Haryana*

- Worked on designing and developing scalable ETL pipelines using Python, PySpark, and SQL to ingest, transform, and load large volumes of structured and semi-structured data.
- Leveraged AWS services such as S3, Lambda, Glue, Athena, and RDS to build cloud-based data engineering solutions.
- Built and optimized PySpark jobs for data processing, achieving significant improvements in job execution time and resource utilization.
- Developed and maintained SQL scripts for data extraction, transformation, and validation across multiple data sources.
- Collaborated with cross-functional teams to understand data requirements and deliver clean, reliable, and timely datasets for analytics and reporting.
- Worked on two major projects involving end-to-end data pipeline implementation, cloud deployment, and performance optimization—gaining valuable hands-on experience with real-world big data solutions.
- Gained strong understanding of data architecture, data modeling, workflow orchestration, and monitoring best practices.

Projects

Point and Click ML | *Python, PySpark, AWS, Flask, Docker, PostgreSQL/PostGIS*   **2024**

- Designed and implemented an automated machine learning system to predict binary outcomes using geospatial and demographic data derived from geographic inputs.
- Built scalable ETL pipelines with PySpark and AWS Glue to process and transform large volumes of location-based data.
- Developed AWS Lambda functions and integrated them with API Gateway to expose predictions via secure, low-latency RESTful APIs.
- Utilized PostgreSQL with PostGIS for spatial data processing, enabling calculation of regional demographics like population density and age distribution.
- Built backend services in Python using Flask to handle geographic inputs, orchestrate API calls, and serve machine learning predictions.
- Containerized the application using Docker and deployed it on AWS EC2 instances for scalability, maintainability, and streamlined DevOps.
- Led end-to-end development—from data ingestion and feature engineering to model deployment—resulting in a robust, production-ready ML solution.

Data Validator | *Databricks, Python, SQL*   **2024**

- Designed and developed a data validation framework to compare and validate records between source and target tables in ETL workflows.
- Implemented logic in Databricks using PySpark and SQL to automate row-level and column-level data integrity checks across large datasets.
- Built a reporting system that generates comprehensive validation reports summarizing successful validations, failed records, and discrepancies with detailed metrics.
- Enabled automatic flagging and logging of mismatched records to facilitate root cause analysis and improve data pipeline reliability.
- Contributed to overall data quality assurance efforts by integrating the validator into end-to-end data workflows.

Technical Skills

Languages: Python, Java, C++, JavaScript, SQL, PostgreSQL, MongoDB, PySpark

Cloud Platforms & ETL Tools: AWS, Azure, Databricks, AWS Glue

Developer Tools: VS Code, Eclipse, Google Cloud Platform, Android Studio, DBeaver

Technologies/Frameworks: Linux, Jenkins, GitHub, Bitbucket, Docker

Leadership / Extracurricular

Tech Community Club   **2023 – Present**

Core Member / Technical Mentor   *Sanatan Dharma College*

- Organized hands-on workshops and knowledge-sharing sessions on Python, SQL, and cloud platforms like AWS and Azure.
- Mentored junior students on building real-world data engineering projects, including ETL pipelines and API integrations.

- Collaborated with team members to host coding competitions and webinars, increasing peer engagement and technical awareness.