Santhosh bharadwaj H A

Mob:+91 8970210996 Email: santhoshbharadwaj2498@gmail.com
https://santhoshbharadwaj-h-a.github.io/portfolio/

Professional Summary:

- **Data Engineer with 4+ years of experience** in designing and implementing high-performance data solutions. Proficient in Python, PySpark, SQL, AWS, and Pandas, with a strong focus on developing scalable data pipelines, optimizing processes, and ensuring data integrity.
- **Skilled in problem-solving** with extensive knowledge of big data technologies and cloud platforms.
- **Successfully built and maintained ETL pipelines** using Python and PySpark to efficiently process and transform large datasets.
- **Enhanced query performance by 20%** through optimizing SQL queries for faster data retrieval and aggregation.
- **Designed and deployed data warehousing solutions on AWS**, enabling advanced analytics and data-driven decision-making.
- **Engineered seamless data integration workflows**, ensuring efficient data movement across various systems and platforms.
- **Performed thorough data quality assessments**, implementing data cleansing and validation techniques to uphold accuracy and reliability.
- **Collaborated with cross-functional teams** to understand business data needs and deliver customized technical solutions supporting data-driven initiatives.
- **Architected and implemented high-performance RESTful APIs using Django and Flask**, incorporating best practices in authentication, pagination, and endpoint optimization to deliver robust and scalable backend services.

Professional Experience:

- **Medilenz Innovation Private Limited**
  Software Engineer — April 2024 – Present
- **SMS TechSoft**
  Data Engineer — February 2021 - February 2024

Technical Skill Set:

- Programming Languages: **Python**
- Big Data Technologies: **PySpark**
- Databases: **SQL, Amazon Redshift, Athena, Postgres**
- Cloud Platforms: **Amazon Web Services (AWS)**
- Data Processing & Analysis: **Pandas, Redshift, Athena**
- ETL Tools: **Apache Airflow, AWS Glue**
- Version Control: **Git**
- Integrated Development Environment (IDE): **Jupyter Notebook**
- API Development: **Django REST Framework, Flask-RESTful**
- NoSQL Databases: **MongoDB, DynamoDB**
- Monitoring Tools: **Prometheus, Grafana**

Education:

- **BE**. in **Information Science** & **Engineering**
  Visvesvaraya Technological University 2017-2021

- **Diploma in computer Application (DCA)**
  Shri Swami Vivekananda institute Graduated: June 2017

Projects:

1.Title	**Data Lake Automation with AWS Glue for Sales optimization**
Description	Develop an automated data lake solution on AWS using Python and services like S3, Glue, and Redshift, enabling seamless data ingestion, processing, and querying. The pipeline ingests data from various data sources, performs transformations, and loads the processed data into Redshift data warehouse for analysis.
Period	April 2024 – Present
Position	Software Engineer
Responsibilities	• Design and configure the data lake architecture • develop ETL workflows using AWS Glue crawler, Glue jobs , handle data ingestion and transformation • Load the transformed data into Redshift for further analysis and reporting. • Ensure data quality and integrity.
Technical Skills	**Python, SQL, AWS, PySpark, AWS Glue, Redshift, Apache Airflow**

2.Title	**Personalized Learning Pathways with Real-Time Feedback.**
Description	develop and maintain an end-to-end data pipeline for a personalized learning platform. This involved extracting, cleaning, and transforming data from various sources, integrating it into a scalable AWS Redshift data warehouse using tools like Apache Airflow and Python to ensure high-quality, real-time data for optimized learning experiences.
Period	February 2021 - February 2024
Position	Data Engineer
Responsibilities	• to extract data from the different data sources • including user interactions, user profiles and feedback. • Build ETL pipelines to integrate and aggregate data from different sources into • a unified data warehouse. Using tools like Apache Airflow, DBT, or custom scripts for this purpose. • Implement data cleaning and transformation processes to ensure data quality. • Use tools like Apache Spark or Pandas for data processing. • Store processed data in a scalable data warehouse (e.g., AWS Redshift) optimized for querying.
Technical Skills	**Python, Pandas, SQL, AWS, Apache Airflow, Redshift**

Declaration
I hereby declare that the information furnished above is complete and true to the best of my knowledge.
Place: Bangalore