

UJJWAL SONTAKKE

Data Engineer

Summary

I'm a goal driven experienced Data Engineer with **2+ years of experience**, passionate about **Big Data Tech Stack**. I am committed to team success through detail-oriented work & also proficient in **handling & extracting insights from large-scale data**.

Work Experience

Software Engineer | HCLTech, Pune, Maharashtra

Feb 2022 – Present | (2+ Years)

Project 1: Credit Card Fraud Detection

I contributed significantly to the successful implementation of a credit card lending service, emphasizing fraud detection and transparency.

Notable tasks include:

- Designed and implemented a scalable **AWS-based** data processing pipeline for handling large transaction data of **3 million** daily record.
- Achieved 20x faster execution with **Apache Spark-Scala** on **AWS EMR**, maximizing throughput and resource utilization through performance tuning techniques.
- Leveraged **Amazon Redshift** for a fast, managed data warehouse, significantly improving **JOIN** query performance via partitioning and bucketing on normalized tables.
- Designed and implemented automated ETL workflows with **AWS Glue**, ensuring data consistency and quality throughout the pipeline.
- Enhanced efficiency and minimized latency by utilizing **incremental load** techniques and optimized file formats (CSV, Avro, Parquet).
- Considered exploring **Amazon Athena** for serverless querying.
- Utilized **Amazon CloudWatch** to proactively monitor performance metrics and ensure continuous optimization.

Project 2: Database management for assisting mortgage underwriter

Team Member | Duration: 6 months

In this project, Utilized SQL to manage extensive datasets supporting mortgage underwriters in income, credit, asset, and property analysis. Employed normalization techniques to denormalized forms, addressing discrepancies in self-occupancy, rental income, credit scores, and employment gaps across multiple credit validations. Implemented normalized tables into denormalized structures for effective underwriting assessments.

Some key achievements include:

- Processed a daily average of **2 million** data records for Loan Documentation alignment, ensuring strict compliance with respective agencies policies.
- Employed **MySQL's** robust features, including constraints like **NOT NULL, UNIQUE, and CHECK** constraints, resulting in a significant reduction in data errors and ensuring data reliability.
- Proficiently utilized Microsoft Excel for precise data alignment, maintaining data accuracy and quality.
- Designed and optimized SQL queries to handle large datasets, including critical operations such as credit checks, income calculations, and debt-to-income ratios, reducing query execution time by a remarkable **10-times**.

This optimization expedited the underwriting process and empowered quicker, well-Informed decisions by the underwriting team.

Contact

Phone: +91 9595282625

Email: ujjwals0511@gmail.com

LinkedIn:

<https://www.linkedin.com/in/ujjwalsontakke511/>

Portfolio

GitHub Repository:

<https://github.com/Sujiwal511/BigDataTechStack>

Skill Sets

Technical Expertise

Big Data: Hadoop, Hive, Apache Spark, PySpark

Database: MySQL,

NoSQL (H-Base)

Language: SQL, HiveQL, Scala, Python

AWS: S3, EMR, Athena, RedShift, Glue

Scheduler: Airflow

CI/CD: Git

Scripting: Shell Scripting

Other Tools:

MS Office, IDE Eclipse,

Cloudera

Education

Bachelor of Engineering (CSE)

Everest College of Engg.

Chh. Sambhajinagar (Abad)

8.08 CGPA | 2018-2021

Certifications

• AWS Certified Trainee: AWS Cloud Practitioner Essentials

• AWS Certified Trainee: Building Data Lakes on AWS (Technical)

• Big Data Engineering Master's Program, TrendyTech, Bengaluru
Feb 2023 – Feb 2024