

EDUCATIONSan Francisco State University, San Francisco, CA
Master of Science in Business Analytics

Jan. 2020 - Dec. 2021

GPA: 3.7

Amravati University, Amravati, MH, India

Jun. 2010 - May 2014

Bachelor of Engineering (B.E.) in Electronics and Telecommunications

GPA: 3.8

KEY HIGHLIGHTS

- Data Warehousing (SQL, PostgreSQL)
- Programming (Python, Spark, Scala)
- Distributed Computing (Docker, Kubernetes)
- Orchestration (Airflow, Step Function)
- Data Streaming (Kafka, Flink, Kinesis)
- Data visualization and storytelling (Tableau)

SKILLS

Programing Languages : Python, PySpark, Scala, Java
Database/Query : PostgreSQL, Hive, MySQL, Oracle, MongoDB, Big Query, Cassandra, Presto
AWS Services : S3, Lambda, Glue, Step Functions, Athena, EMR, Redshift Spectrum, DynamoDB.
Distribution Frameworks : Apache Spark, Hadoop, Kubernetes, Docker
Real-time Streaming : Apache Kafka, Apache Flink, Apache Hudi, Delta Lake

WORK EXPERIENCE**Amazon.com - Data Engineer**, San Francisco, CA

Jun. 2021 – Present

Transactional solution to spark architecture and behavior using Delta Lake - [Link](#) (Delta lake, Python, SparkSQL, Scala)

- Designed and implemented a data pipeline solution that improved data processing speed by ~30% and reduced errors by 83%, resulting in significant cost savings and increased operational efficiency for the organization
- Dynamic file compaction using Delta Lake - [Link](#)
- Orchestration of new partition synchronization from AWS S3 to Redshift - [Link](#)

Row level security and Redshift audit mechanism

(Python, SparkSQL, Tableau, AWS Services)

- Developed and implemented a GDPR-compliant data engineering workflow for EU employee data governance using Airflow, S3, SQS, Glue, Redshift, and Tableau.
- Implemented robust row-level security for leadership onboarding dashboard to restrict access to sensitive EU employee data, enhancing data security and compliance and reducing the risk of costly GDPR violations around 10M Euros.
- Achieved operational excellence by performing data quality checks, auditing mechanisms, anomaly detection

Revolutionizing Headcount Reporting: One-Stop Solution for Metrics Management - [Link](#)

(Spark, AWS Services, Tableau)

- Architected and implemented a serverless data pipeline using AWS Glue, AWS Lambda, and Amazon S3 to extract, transform and load headcount data from various sources, enabling seamless integration and data governance across multiple line of businesses.
- Designed and implemented a scalable, fault-tolerant data warehousing solution on Amazon Redshift, leveraging AWS services for an IQR-based anomaly detection framework to ensure high-quality and consistency of data
- Developed a Tableau dashboard, providing valuable insights into data & saving 15 hrs. per department on generating BvA reports

Consumer Engagement Funnel Analysis and User Acquisition Optimization - [Link](#)

(PySpark, SQL, Tableau)

- Designed and implemented a comprehensive framework to analyze user behavior across different stages of the consumer engagement funnel, from Signup to First order, soft/hard churn, and resurrection, resulting in a 20% increase in user engagement and a 15% increase in customer retention rate.
- Developed a scalable PySpark model to analyze user acquisition funnel leakage, resulting in a 10% reduction in user acquisition cost and an 8% increase in customer lifetime value, and created a user-friendly Tableau dashboard for improved decision-making

Improving LTV Efficiency: Maximizing Revenue with Customer Acquisition Cost(CaC) Analysis - [Link](#) (Python, SQL, AWS Services)

- Developed an LTV Efficiency Dashboard using Amazon Redshift, S3, and QuickSight to visualize and analyze customer lifetime value by customer acquisition cost, optimizing marketing spend for revenue growth.
- Analyzed customer acquisition cost and lifetime value data using SQL on Redshift and machine learning on SageMaker, resulting in 15% lower acquisition cost and 12% higher lifetime value.
- Automated data processing & visualization with Python & AWS Lambda/CloudFormation, increasing speed by 20% & streamlining dashboard updates.

IBM - Senior Data Analyst, Pune, MH, India.

Aug. 2017 – Feb. 2019

- Developed and maintained data marts using SQL and ETL tools such as Informatica, enabling stakeholders to access key business metrics in real time and increasing data accuracy by 98%.
- Created customized reports using Python and Tableau, providing insights into customer behavior, market trends, and business performance, resulting in a 6% increase in revenue.

Syntel - American Express – Senior Data Analyst, Pune, MH, India

Oct. 2014 – Aug. 2017

- Developed SQL scripts (MySQL) to categorize the legacy data into various data marts created as per business requirements.
- Redesigned SQL stored procedures through ETL Informatica workflows, reducing database utilization by 13%
- Created data model consisting of various data marts to hold enterprise data in standardized schema using ETL Informatica Cloud

Certifications

- AWS Certified Big Data- Specialty
- Google Cloud Professional - DE
- Hackerrank Gold Badge – Python and SQL