

EDUCATIONSan Francisco State University, San Francisco, CA

Jan. 2020 - Dec. 2021

Master of Science in Business Analytics

GPA: 3.8

Courses: Time series Forecasting, Business Intelligence, Data Analysis, Machine learning, NLP, Statistics

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)
- Data viz. and storytelling (Tableau, Looker)
- Orchestration (Airflow, AWS Step Func)
- Machine Learning and prediction models
- Experimentation and Hypothesis testing

SKILLS**Programming** : Python, Apache Spark, Scala, Linux**Database/Query** : PostgreSQL, Hive, MySQL, Oracle, MongoDB, Big Query, Cassandra, Presto, InfluxDB, Clickhouse DB**Visualization Tools** : Tableau, Looker, Google Studio, QuickSight, Power BI**Tools** : Apache Airflow, Docker, Git and GitHub, FiveTran integration tool**AWS Cloud** : S3, Lambda, Glue, Step Functions, Athena, EMR, Redshift Spectrum, DynamoDB.**Google Cloud Platform** : BigQuery, Cloud, storage, cloud function, pub/sub, Dataflow, Dataproc, Cloud Data Fusion, Spanner DBWORK EXPERIENCE**Exabeam Inc. – Data Engineer II**, Foster city, CA

July. 2023 – Present

Automated Optimization of BigQuery Table Partitioning and Clustering - [Link](#) (BigQuery, GCP Functions, cloud Composer)

- Developed and deployed a workflow using Google Cloud Composer (Apache Airflow) to automate the optimization of BigQuery table partitioning and clustering, enhancing query performance by up to 70% and reducing data storage costs.

FinTech wide Data Quality Framework - [Link](#)

(Airflow, BigQuery, Python, Dataflow, Pub/Sub)

- Designed and developed a scalable Fintech anomaly detection framework (FTDQ) that increased true positive accuracy by 50% and reduced false positives by 55%, saving 8+ engineering hours per week.
- Led the migration of 100+ jobs from legacy systems to the FTDQ framework, onboarded 70+ anomaly detection checks, and facilitated the adoption by 8 teams across the organization.

Amazon.com – Data Engineer, San Francisco, CA

Jun. 2021 – July 2023

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)

- 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

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 Service)

- 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.

Oct. 2014 – Feb. 2019

- Developed and maintained data marts using SQL and ETL tools such as Informatica, enabling stakeholders to access key business metrics 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.

Certifications

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