YOGESH J. BORICHA

College Park, MD | +1 (240)-927-8677 | https://www.linkedin.com/in/yogeshboricha/ | yogeshboricha6@gmail.com

EDUCATION

University of Maryland, College Park, MD — July 2022 - Present
Masters in Information Management — GPA - 3.8
Courses: Big Data Infrastructure, Data Visualization, Data Analytics for Information Professionals, Natural Language Processing (NLP), Information Technology and Organizational Context, Database Design, Management Concepts and Principles for Information Professionals, Data Integration, Users and Use Context, UX Strategy

University of Mumbai, Mumbai — August 2018 - June 2022
Bachelors of Technology — CGPA - 7.3
Courses: Statistics, C Programming, Data Analytics, Product Development Basic Electrical Engineering, Economics and Finance, Project Management

TECHNICAL SKILLS

Programming Languages: SQL, R, Python (NumPy, Pandas, Scikit Learn, Matplotlib, PyTorch, TensorFlow)
Databases: Oracle Database, Microsoft SQL Server, MySQL, PostgreSQL, T- SQL, MongoDB, Amazon DynamoDB, Amazon Redshift
Extract, Transform and Load (ETL): AWS Glue, Apache Airflow, Microsoft SQL Server Integration Services (SSIS), Apache Spark, Analysis Services (SSAS)
Data Visualization: Tableau, Power BI, Google Data Studio (Looker Studio), SAS Visual Analytics
Tools and Services: Apache Airflow, Cradle, Jupyter Notebook, Android Studio, Git, Google Cloud Platform (BigQuery, Airflow, Dataflow, Dataproc), Microsoft Azure (Azure Data Factory, Azure SQL Database), Big Data Technologies (Hadoop, Apache Hive), Snowflake, Amazon Web Services (S3, EMR, Kinesis, Firehose, Lambda)

EXPERIENCE

Graduate Teaching Assistant, University of Maryland (iSchool), College Park, MD — August 2022 - Present
- Pioneered a vital role in teaching fundamental concepts of databases, E-R diagrams, SQL, normalization, MySQL and NO-SQL databases to a diverse group of 80+ undergraduate students in the course "Database Design and Modelling"
- Hosted successful weekly office hours integrating advanced visualization tools such as Tableau to showcase practical examples and illustrate complex database structures visually compelling manner
- Evaluated and graded online/in-class assignments and activities presented by students, applying Excel for data analysis and generating comprehensive reports to track individual progress and provide personalized feedback

Data Engineer Intern, Biocon — September 2021 - March 2022
- Implemented a dynamic Power BI dashboard for the production team, targeting a 15% reduction in average open work order resolution time
- Orchestrated complex business intelligence reports in Power BI, integrating data from Snowflake databases, to offer comprehensive insights
- Devised data mining operations within SQL to uncover trends, patterns, and critical performance metrics within intricate datasets and collaborated with the team to seamlessly integrate analytical insights into Power BI, ensuring precise data aggregation and analysis
- Developed and managed real-time data streaming solutions using AWS Kinesis and Firehose, enhancing data ingestion and processing capabilities
- Evaluated and implemented Apache Airflow on Azure to streamline data pipelines, leading to improved data processing efficiency within the BioPharma domain
- Designed and implemented ETL workflows in Azure Data Factory to ingest 4 terabytes of data per month from 5 different sources into the Azure cloud platform
- Optimized database performance by 20% through indexing strategies and query optimization techniques in MySQL and developed data quality checks and validation procedures, reducing error rates by 15%
- Leveraged AWS Lambda for serverless computing and automated tasks, improving operational efficiency and scalability

Data Engineering Intern, Deloitte — May 2021 - July 2021
- Spearheaded comprehensive market research and analysis, generating data-driven insights into client industries, competitors, and market trends
- Evaluated large volumes of data by cleaning and organizing 100,000+ copies of abnormal data and created a comprehensive brand database by employing Tableau by increasing data accuracy by 25%, and reduced data analysis time by 40%
- Utilized ETL tools like SSIS to ingest data from various sources into data platforms, ensuring seamless data integration and accessibility for analysis
- Employed MySQL to create and execute queries, extracting specific information and generating customized reports for data analysis and decision-making
- Automated data ingestion pipelines using Apache Airflow to schedule and orchestrate data movement from source systems to data warehouse at hourly intervals, improving data freshness by 60%
- Designed and performed data quality checks using Python library Pandas to identify and rectify inconsistencies in incoming data streams, reducing data errors by 30%

PROJECTS

Data Architecture Project, Maryland Governor's Office of Community Initiatives (GOCI), MD — January 2024 - Present
- Orchestrated end-to-end data architecture activities encompassing data ingestion, integration, retrieval, storage, and pipeline development, ensuring seamless data flow
- Utilized Amazon S3 for cost-effective storage of infrequently accessed data, backups, and archives, resulting in a 50% decrease in storage expenses
- Leveraged Amazon EBS for high-performance storage of OLTP databases and data warehouses, enhancing data retrieval accuracy by 95%
- Designed and accomplished robust data models and schemas to maintain data consistency, integrity, and security, adhering to best practices in data governance
- Developed customized data visualization dashboards using Tableau to empower stakeholders with intuitive and interactive insights into community metrics, fostering informed decision-making

App Development Project, Water Emergency Team, University of Maryland, MD — August 2023 - January 2024
- Developed a high-performance mobile application using Android Studio, Java, and React to empower users to report sanitary sewer overflows (SSOs) to the Water Emergency Team, resulting in a 20% increase in reported SSOs
- Leveraged the Android SDK to access device features, achieving a 30% reduction in average SSO report submission time
- Established a robust and scalable data architecture leveraging JSON and Firebase Realtime Database to efficiently manage and analyze high volumes of Sanitary Sewer Overflow (SSO) reports
- Utilized Git version control for managing codebase changes, fostering seamless collaboration with a team of 3 developers, and tracking project history, ensuring code integrity and efficient project management
- Integrated Firebase Realtime Database to store and manage SSO reports submitted by users, ensuring data persistence, security, and real-time synchronization across devices

US Elections Data Analysis: — May 2020 – April 2021
- Performed a comprehensive exploratory data analysis (EDA) on US Elections data using SQL queries within MySQL Workbench to identify trends, patterns, and anomalies with 25% improvement in data accuracy
- Leveraged SQL queries to extract and analyze relevant product performance data from databases, enabling data-driven decisions that resulted in a 30 % improvement
- Designed and deployed data pipelines using Apache Spark for distributed data processing of US Elections data, resulting in a 30% improvement in processing speed
- Integrated BigQuery for scalable and efficient data processing, further optimizing performance and scalability