

SURBHI DILIP BHOR

+1 (480) 761-9742 | CHAMPAIGN, IL | SDBHOR2@ILLINOIS.EDU | LINKEDIN

EXPERIENCE

Data Engineer Intern | Boston Consulting Group (BCG), New York June 2023 – Aug 2023

- Built **Azure Synapse data pipelines** to initiate ETL process for efficient initial and incremental data loads across data layers
- Transformed and curated over **25000** records into a staging **Hive** table using **ADLS** and data warehouse resulting in a **SCD1 delta** table, reducing data discrepancies by **75%**
- Implemented data quality validation with **PySpark notebook**, resulting in the identification and resolution of a critical bug, enhancing data integrity
- Strategically assessed six **ERD** migration tools, seamlessly transitioning to the optimal solution, resulting in enhanced real-time collaboration and significant time-saving advantages
- Orchestrated comprehensive **data modelling** for a complex domain spanning 15 diverse tables, translating it into an effective ER diagram and further enhancing understanding through strategic creation of **data flow diagrams** for multiple pods

Senior Analyst (Automation Engineer) | Capgemini, India Oct 2021 – Jun 2022

- Automated critical AWS functionalities like **CloudWatch logs** and **DynamoDB** using **JavaScript** to boost test coverage to **70%**
- Developed interactive dashboards in **Azure DevOps** and **Microsoft PowerBI** for tracking build failures, reducing recurring failures by **21%**
- Performed **data-driven** testing using test data stored in a JSON file to build both positive and negative scenarios into a single test
- Leveraged AWS framework to integrate the smoke, regression, integration suites into the **Azure DevOps** and **Jenkins CI/CD** pipeline by scripting over **200** testcases
- Devised **160** and automated **80** integration scenarios, cutting the integration cycle by **40%** and **saving 2 days**

Analyst | Capgemini, India Sept 2020 – Sept 2021

- Executed test cases on **Postman** and analysed the results by creating logs to document testing phases and defects
- Collaborated with **business and engineering teams** to ensure quality throughout the software development lifecycle
- Reported **5%** of the high priority bugs and **led two pods** to ensure defect-free modules during User Acceptance Testing (UAT)
- Collaborated closely with cross-functional internal and client teams using **agile** methodology to ensure quality throughout the software development lifecycle
- Trained a batch of **20 new resources on 3 different areas**, along with training documents and video guidelines for reference

PROJECTS

Instacart Market Basket Analysis | Tableau, Python May 2023

- Applied **dimensional modeling** techniques, including star schema-based data models, **ER** diagram and designed a Tableau **dashboard** to analyze customer behavior and product performance, providing valuable insights
- Conducted exploratory data analysis using Python libraries such as **NumPy** and **Pandas**, leveraging **seaborn** and **matplotlib** for visualizations

Crimes against women | SQL, PowerBI May 2022

- Populated crimes against women database with over **10,000 rows** and **11 columns** ensuring **ACID** properties to maintain integrity and consistency of data using **SQL**, which reduced data discrepancy by **11%**
- Created an interactive dashboard in **PowerBI** to examine historical crime trends and demonstrated how crimes against women increased exponentially from 2001 to 2014, in which approximately **5M** crimes were reported

Flight Booking System | MEAN Stack Nov 2020

- Developed a full-stack application using Node.js and Express.js, serving a **REST API** with Angular as the frontend. Implemented **micro-services**, JWT authentication, payment gateway (Razorpay), email (nodemailer), API documentation (Swagger Hub), testing (Mocha & Chai)
- Deployed application on **AWS** server using lambda functions and S3 bucket along with the migration of the entire database from MongoDB to DynamoDB using DMS-Database Migration Service

SKILLS

Programming Languages: Python, SQL, JavaScript

Libraries and Frameworks: PySpark, Angular.js, Node.js, Express.js, Chai assertion, Jenkins, Swagger Hub

Cloud: AWS - Cloud9, Lambda functions, S3 bucket, CloudWatch logs, DynamoDB (NoSQL), API Gateway

Azure - Synapse Analytics, Data Lake Storage Gen2, Blob Storage, DevOps

Tools and IDE's: VSCode, MySQL Workbench, R Studio, Jupyter Notebook, Git, Postman

Database and Analytics: Tableau, PowerBI, MS Excel, MySQL, MongoDB (NoSQL)

APIs and Protocols: RESTful APIs, JSON, HTTP/HTTPs

EDUCATION

University of Illinois at Urbana Champaign Aug 2022 – May 2024

Master of Science in Information Management

CGPA: 4.0/4.0

University of Mumbai

Bachelor of Engineering in Electronics Engineering

Aug 2016 – Oct 2020

CGPA: 8.01/10.0

LEADERSHIP

- **Course Assistant:** MCS program – Foundations of data curation, University of Illinois
- **Pod Lead:** Spearheaded quality assurance (QA) of 2 micro-services in Capgemini
- **Dance Head:** Led a dance crew of 20 people in intercollegiate and national competitions
- **Marketing Head:** Handled social media and marketing initiatives for IEEE student branch