Mihir Dhirajlal Satra
Location: San Jose, CA
Contact: +1 408-394-6683
Email: mihirdhirajlalsatra@gmail.com

Education:
Master of Science in Computer Science – San Jose State University (August 2023 – May 2025):
GPA: 3.83/4.0: This represents an excellent academic standing, indicating high performance in coursework and a thorough understanding of complex concepts in computer science.
Coursework:
Distributed Computing: Focuses on systems where components located on different networked computers communicate and coordinate to achieve a common goal. Explored algorithms for distributed systems, issues like fault tolerance, and the challenges of concurrency.
Database Systems: Studied the theory and application of database systems, focusing on relational databases, query optimization, normalization, indexing, transactions, and consistency models.
Big Data: Investigated the methods for processing vast datasets, including distributed file systems like HDFS, frameworks like Hadoop and Spark, and the challenges posed by high-volume data processing.
Cloud Computing: Learned the infrastructure and services provided by cloud platforms like AWS and Azure, including concepts like virtualization, scalability, and the use of cloud resources for application deployment.
NoSQL Databases: Examined databases that are non-relational, focusing on systems like MongoDB and Cassandra, and when to use NoSQL solutions over traditional relational databases.
Artificial Intelligence: Explored machine learning algorithms, neural networks, reinforcement learning, and other AI techniques to solve problems such as classification, regression, and decision-making.
Bachelor of Science in Computer Engineering – University of Mumbai (August 2017 – May 2021):
Coursework:
Data Structures and Algorithms: Gained deep knowledge of data structures (like arrays, linked lists, trees, graphs) and algorithms (sorting, searching, dynamic programming, etc.), essential for problem-solving and optimization in computer science.
DBMS (Database Management Systems): Studied the architecture of databases, how to manage data efficiently, and how to design databases that balance performance, security, and scalability.
Operating Systems: Learned the principles of modern operating systems, including process management, memory management, file systems, and I/O.
Software Engineering: Focused on the structured process of developing software, from requirements gathering and analysis to design, implementation, testing, and maintenance.
Web Technology: Acquired knowledge of web development technologies (HTML, CSS, JavaScript) and back-end technologies for server-side processing.
Cloud Computing: Introduced to cloud environments, their underlying infrastructure, and the services available for hosting scalable, distributed applications.
Skills:
Programming: Proficient in multiple languages across various paradigms:

Python, Golang, Java: Familiarity with high-level, object-oriented languages for general-purpose programming, with applications ranging from web development to data science and backend engineering.
SQL: Expertise in writing structured queries to retrieve and manipulate data stored in relational databases.
JavaScript: In-depth knowledge of client-side scripting for web applications.
TypeScript, C, C++, Rust: Understanding of strongly typed languages for performance-critical applications, such as system programming.
Scala, C#: Experience with languages that are prominent in enterprise-level applications and functional programming.
HTML, CSS: Foundational web development skills to structure and style web pages.
DAX (Data Analysis Expressions): Proficiency in creating advanced business logic for Power BI and Excel, used to perform complex calculations on data models.
Frontend/Backend Development: Experience with various frameworks for developing web and backend services:

NodeJS, AngularJS, ExpressJS: Understanding of JavaScript-based environments for building scalable server-side and front-end applications.
ReactJS, Vue.js: Skills in modern JavaScript frameworks for building dynamic user interfaces.
Django, Flask: Experience with Python frameworks to create efficient web applications and REST APIs.
Springboot: Knowledge of Java-based framework for building microservices and enterprise-level applications.
CI/CD, Git: Familiar with continuous integration and deployment pipelines, automating the testing and deployment of applications through tools like Jenkins, CircleCI, and version control systems like Git.
Industry Knowledge: Expertise across various domains:

Data Warehousing and ETL (Extract, Transform, Load): Proficient in techniques for data integration from multiple sources, transforming it into a format suitable for analysis, and loading it into data warehouses.
Big Data: Familiar with large-scale data processing tools (like Hadoop, Spark) to handle datasets that exceed the capabilities of traditional databases.
Full Stack Development: Able to build both client-facing applications and server-side logic.
Machine Learning and NLP: Knowledgeable in machine learning algorithms and natural language processing techniques, applying them to problems like classification and sentiment analysis.
Cloud Technologies:

Azure: Worked with Data Lake, Synapse Analytics, Databricks, and Analysis Services to design and implement cloud-based data solutions.
AWS: Experience with S3 (cloud storage), Redshift (data warehousing), DynamoDB (NoSQL database), EC2 (cloud computing), and Lambda (serverless computing).
Data Technologies:

Hadoop, Spark: Familiar with distributed computing platforms for big data processing.
Kafka, Docker, Kubernetes: Knowledge of event streaming platforms and containerization tools for deploying applications.
Snowflake, Airflow: Experience in cloud-native data warehousing and workflow orchestration tools for automating tasks.
BigQuery, Grafana, Tableau: Proficiency with cloud-based analytical databases and visualization tools.
Databases: In-depth understanding of various databases:

Relational: MySQL, PostgreSQL for structured data storage and querying.
NoSQL: MongoDB, Cassandra, Redis for handling unstructured or semi-structured data.
Search Engines: Elasticsearch for efficient search and analysis.
Distributed Databases: HBase, Hive, Google Spanner, and BigTable for handling large-scale, geographically distributed data.
Work Experience:
Software Engineer Intern – Nutanix (May 2024 – Present):
Developed Distributed Tracing using Jaeger and OpenTelemetry: Built a tracing system using Jaeger (an open-source, end-to-end distributed tracing tool) and OpenTelemetry, which helps in tracking system requests and responses across microservices. This provided real-time visibility into how requests flowed through services in a distributed architecture, resulting in a 40% improvement in system visibility and operational efficiency.
Service Dependencies Analysis in Microservices Architecture: Conducted detailed analysis of interdependencies between services in a microservices-based architecture. This helped visualize the flow of data between services, identifying inefficiencies and reducing the complexity of the system.
Elasticsearch Querying and Indexing Optimization: Enhanced Elasticsearch's ability to process and retrieve data by optimizing query patterns and indexing strategies, achieving a 40% improvement in performance for data retrieval tasks.
Performance Improvement through Latency Reduction: By analyzing trace data, identified bottlenecks causing latency issues. Implemented improvements that reduced the time it took for services to respond, enhancing overall system performance by 30%.
Senior Data Engineer – LTIMindtree (July 2021 – July 2023):
Azure Synapse Data Warehouse Development: Led the design and deployment of a data warehouse on Azure Synapse for a life sciences company. This architecture consolidated vast amounts of data into a centralized location, cutting storage costs by 20% and improving data processing efficiency by 40%.
Revenue Generation via Production Releases: Managed multiple production releases using Azure DevOps, which led to $36 million in revenue. These releases included crucial features for a manufacturing analytics system that needed to meet tight deadlines.
Automation with Azure Databricks: Automated the data transformation process using Python notebooks on Azure Databricks. This reduced the manual data cleaning effort by 25%, freeing up time for more complex data analysis.
Azure Analysis Services Tabular Models: Created efficient tabular models with embedded business logic, reducing the load time for Business

Intelligence (BI) reports by 85%, which greatly sped up data access for decision-makers.
Power BI Report Development: Designed over 50 Power BI reports, streamlining reporting across the organization, which increased the overall efficiency of decision-making processes by 15%.
Software Engineer Intern – GRT Global Logistics Pvt. Ltd. (December 2019 – January 2020):
Selenium Automation for ERP Software Testing: Developed automated test scripts using Selenium, a widely-used web testing tool, to automate testing for the company's ERP system. This decreased the time required for testing by 80%.
Increased Bug Resolution Efficiency: Automated testing allowed for quicker identification and resolution of system bugs, which improved the overall quality of the software and reduced downtime by 15%.
System Design Enhancement: Collaborated with the software team to improve the system's overall design, boosting the operational efficiency of the ERP platform by 25%.
Projects:
Personality Prediction System to Improve Employee Recruitment:
Developed a system to assist HR departments in predicting candidate personalities based on questionnaire responses. Used machine learning models such as Google BERT (a transformer-based model for natural language processing tasks) to extract keywords from responses. Additionally, logistic regression models were used to predict personality traits based on the Big Five Personality Model.
The project also employed fuzzy logic to assess a spectrum of personality characteristics, making the system flexible and adaptable to various recruitment contexts.
This project was published in Springer Nature under 'Emerging Applications and Strategies' and indexed by the Scopus Database.
Ride Insights: Analyzing NYC Taxi Trip Data:
Built an end-to-end data analytics pipeline using Google Cloud Platform (GCP) services. The project processed NYC taxi trip data and provided actionable insights for city planning and transportation services.
Implemented an ETL (Extract, Transform, Load) process using Mage, which streamlined the processing of data from Cloud Storage (GCS) to BigQuery, reducing data processing time by 30%.
Created dashboards in Looker Studio, making the processed data more accessible to stakeholders and improving user engagement through interactive data visualizations.