

Darshan Nilesh Mahajan

AWS Certified Solutions Architect Associate

darshann@usc.edu | +1(323)2053739

linkedin.com/in/darshan-mahajan/ | github.com/i4mdarshan

i4mdarshan.github.io/portfolio/

PROFESSIONAL SUMMARY

Software engineer devoted to customers and ownership-driven with, **3 years of experience** taking complex large-scale systems from design to deployment. Proven ability to deliver results by building and optimizing scalable backend services, data pipelines, and REST APIs in fast-paced, agile environments. Expertise in Java, Python, C++, SQL, and AWS.

EDUCATION

University of Southern California

Masters of Science, Applied Data Science (GPA 3.5)

August 2025 – May 2027

Los Angeles, California

- **Relevant Coursework:** Big Data Systems, Distributed Systems, Applied Machine Learning, Knowledge Graphs

University of Mumbai

Bachelors of Science, Electronics & Telecommunications Engineering (GPA 3.6)

August 2018 – June 2022

Mumbai, India

- **Relevant Coursework:** Operating Systems, Database Management Systems, Data Visualization, Data Structures, Algorithms

SKILLS

Cloud & Analytical Tools: AWS (EC2, S3, Lambda, Glue, IAM), PowerBI, MS Excel

Databases: MySQL, MongoDB, Firebase, Hadoop (HDFS, MapReduce)

Programming Languages: Java, C++, Python, PHP, JavaScript, jQuery, TypeScript

Frameworks/Libraries: Pandas, NumPy, PySpark, TensorFlow, Scikit-learn, D3JS, ReactJS, SpringBoot, Laravel

EXPERIENCE

Accenture

November 2022 – August 2025

Mumbai, India

Software Engineer

- Engineered and optimized large-scale **backend REST microservices** and SQL data pipelines processing €10M+ in monthly revenue; applied advanced query optimization and pre-aggregation techniques to reduce report generation time by 40%, directly improving performance for 13 global business units.
- **Designed and developed** a secure, high-availability API component to manage and integrate data flows between distributed enterprise systems (Salesforce, Oracle), ensuring 100% data integrity and compliance for the core billing data platform.
- Contributed to **performance engineering** by optimizing large-scale SQL data pipelines, reducing **query processing time by 40%** through advanced query optimization and pre-aggregation.
- Engineered highly-available Distributed Systems by implementing **master-slave database replication** within **Docker containers**, achieving 99.9% data consistency for the core billing platform.
- Streamlined production Debugging and deployment processes by **building Jenkins** pipelines with Docker support, automating build/deploy cycles and reducing deployment failure rates by 60%.
- **Developed Automation/Scripting tools** (Bash/Shell) to manage legacy system configurations and automatically generate sanitized test datasets, reducing environment setup time by 4 hours per UAT cycle.

DotMinds

April 2020 – April 2022

Mumbai, India

FullStack Developer Intern

- **Applied** solid database fundamentals to design performance-optimized and normalized MySQL schemas for 5+ client applications; **implemented** relational constraints, indexing strategies, and query optimizations to ensure high data integrity.
- **Designed and built** the backend REST APIs and database architecture for a billing automation system (PHP, MySQL) to manage 5,000+ members, successfully automating invoicing and room allocation for over 850 families.

PROJECTS

Distributed Social Graph Analysis

[Python, PySpark, GraphFrames]

- Developed a distributed social graph analytics pipeline on SNAP dataset using PySpark and GraphFrames to compute PageRank, Triangle counts, and Connected Components. Optimized Spark jobs through data partitioning, reducing runtime by 40% and enabling scalable “People You May Know” recommendations.

FrameQL - custom data manipulation library

[Python, Flask]

- Built a custom Python DataFrame library using generators and chunked CSV parsing for memory-efficient, streaming data processing. Implemented a multi-stage query pipeline (filter, join, aggregate) with hash-based joins and grouping logic, enabling GB-scale data analysis without full in-memory loading.

Enterprise Medallion Data Warehouse

[Python, Shell Scripting, Azure, T-SQL]

- Orchestrated a full, end-to-end ETL pipeline for a personal data warehouse by ingesting raw data from Azure Blob Storage, cleaning and transforming it in SQL Server, and modeling it into fact and dimension tables for analytics.

Repeat Buyer Prediction for E-Commerce

[Python, Scikit-learn, TensorFlow, Pandas]

- Increased repeat-buyer prediction accuracy by 10% (validation AUC) for an e-commerce platform by engineering predictive models (XGboost, Logistic Regression) and implementing advanced feature extraction from 6 months of user logs (40,000+ buyers).

Customer Relationship Management (CRM)

[Laravel, Javascript, MySQL]

- Enhanced marketing personalization for a 50K+ customer base by developing a CRM (Laravel, MySQL) with an integrated recommendation engine (collaborative filtering) and a K-Means clustering module, which enabled targeted product suggestions and proactive risk identification.