Thrivikram Kotharu

Thri vi kr am k o tharu@gmai l.com | Mount P leasant, Michi gan | +1 (98 9) 93 3-865 0 | Portf olio | Lin k din

---

Software Engineer with 3+ years of experience designing and building scalable backend systems and data-driven services using Python, SQL, AWS, and Apache Spark (Databricks). Strong background in developing RESTful APIs, cloud-native applications, and automated data processing workflows for high-volume production environments. Proven ability to implement robust testing frameworks using PyTest and Selenium, optimize SQL queries and Spark jobs for performance, and deploy reliable services on AWS (EC2, Lambda, S3, RDS) with monitoring via CloudWatch. Experienced in delivering maintainable, well-tested systems that improve application stability, reduce processing time, and support data-intensive business applications.

Skills

Programming & Tools

• Python, Advanced SQL (CTEs, Window Functions, Query Optimization), Git, GitHub, Linux, Bash (Basic Scripting), Postman, Docker

Backend & API Development

• REST API Development (Flask, FastAPI), Object-Oriented Programming (OOP), Debugging, Performance Optimization

Data Engineering

• ETL Pipeline Development, Data Modeling, Data Quality Frameworks, Analytical Schema Design (Star & Snowflake), Pipeline Validation Testing, Distributed Data Processing

Big Data & Analytics Platforms

• Apache Spark (PySpark), Databricks

Cloud & Databases

• AWS (EC2, Lambda, S3, RDS, CloudWatch), PostgreSQL, MySQL

Testing & Automation

• PyTest, unittest, Selenium (UI & End-to-End Automation), Test Case Design, Regression Testing, Integration Testing

Monitoring & Optimization

• Logging & Monitoring, Performance Tuning, Cost Optimization

Development Practices & Collaboration

• Agile / Scrum, CI/CD Pipelines, CI Basics, Code Reviews, Technical Mentoring, Sprint Planning, Cross-team Collaboration

WORK EXPERIENCE

Senior Software Engineer
Oct 2022 – Aug 2024
Tata Consultancy Services (TCS)

- Led development of cloud-native Python services deployed on AWS (EC2, Lambda, S3, RDS), supporting analytics pipelines processing 1.5M+ records/day.
- Designed analytical data models and wrote advanced SQL queries (CTEs, window functions, aggregations) for reporting and downstream analytics teams.
- Built scalable ETL workflows using Python + Apache Spark (Databricks), reducing batch processing time from 2.5 hours to 35 minutes.
- Developed automated data quality checks and pipeline validation frameworks using PySpark and SQL, reducing data inconsistencies by 60%.
- Created Selenium-based end-to-end test automation for data-driven web applications and dashboards, increasing release confidence and reducing production defects by 38%.
- Authored 120+ automated test scripts (unit, integration, and UI), achieving 88% test coverage across core services.
- Implemented monitoring and structured logging using AWS CloudWatch, cutting production debugging time by 40%.
- Optimized Spark jobs using partitioning and query tuning, lowering AWS compute costs by 22%.
- Mentored 2 junior developers on Python best practices, SQL optimization, testing strategies, and cloud deployment workflows.

Junior Software Engineer, Oct 2021 – Sep 2022 Tata Consultancy Services (TCS)

- Developed and maintained 12+ backend services using Python (Flask & FastAPI) to support internal business applications and REST APIs serving 20K+ daily requests.
- Designed and optimized complex SQL queries, joins, and indexes on PostgreSQL and MySQL databases, improving report generation performance by 32%.
- Built reusable Python modules for authentication, logging, and error handling, reducing duplicated code by 35% across projects.
- Implemented SQL-based data access layers and schema changes to support new features while maintaining backward compatibility.
- Developed UI automation test suites using Selenium (Python) for core workflows, reducing manual regression testing effort by 45%.
- Wrote unit and integration tests using PyTest and unittest, increasing code coverage from 52% to 81% within one year.
- Automated data validation and file processing workflows using Python + SQL scripts, cutting manual operations by 40%.
- Participated in Agile development (Scrum), contributing to 25+ production releases with consistent on-time delivery.
- Collaborated with QA and DevOps teams to debug SQL performance bottlenecks and production issues, reducing incident resolution time by 30%.

RESEARCH EXPERIENCE

Research Assistant (Volunteer) 2025-Present Central Michigan University, Mount Pleasant, MI

- Conducted academic research on platform governance and gig worker benefits, applying a socio-technical systems and algorithmic management perspective.
- Co-authored a research paper accepted and presented at ICIS 2025 TREO (Association for Information Systems).
- Performed comparative analysis of U.S. and Indian gig-economy frameworks, examining regulatory structures, platform design, and digital infrastructure.
- Contributed to literature review, research synthesis, and theoretical framing, translating findings into policy- and platform-relevant insights.
- Performed comparative analysis of **U.S. and Indian gig-economy frameworks**, contributing to literature review, synthesis, and policyrelevant insights.

PUBLICATIONS

Platform Governance and Gig Worker Benefits: A Socio-Technical Approach Kata, L. B. S., Kotharu, T., Huang, X.
ICIS 2025 TREO, Association for Information Systems (AIS), 2025

Project

Instamart Data Warehouse & Dimensional Modeling | Snowflake, AWS S3, SQL | Documentation

- Designed and implemented a Star Schema with 1 fact and 5+ dimension tables, enabling queries across orders, products, and customers.
- Built ELT pipelines from AWS S3 → Snowflake, loading and transforming 3M+ records, improving query performance by 40%.
- Delivered 7+ analytical use cases (basket size trends, reorder rates, peak hours), supporting portfolio-style reporting and insights..

Spotify Playlist ETL Pipeline – Snowflake and AWS | GIT

- Built a serverless ETL pipeline using **AWS Lambda, Glue, S3, and Athena** to ingest and analyze Spotify API data.
- Reduced schema drift issues by **40%** through automated schema validation and cataloging.
- Enabled ad-hoc analytical queries on millions of records with sub-second performance using Athena.

Real-Time Data Streaming Pipeline Using AWS, Apache NiFi, and Snowflake | GIT

- Designed and implemented an end-to-end real-time data pipeline using Apache NiFi, Kafka, Snowflake, and AWS S3.
- Enabled ingestion and processing of 10,000+ records per batch with near-real-time latency.
- Implemented Snowpipe + event-driven ingestion for fully automated data loading.
- Built SCD Type 1 models using Snowflake MERGE statements to ensure consistent analytical views.
- Deployed containerized ingestion services using Docker on AWS EC2, reducing environment setup time by 60%.

Watt's Next: EVs & Clean Energy Impact Tableau Dashboard | Dashboard | Video

- Designed a Tableau dashboard analyzing EV adoption, renewable energy production, CO_2 emissions, and GDP across top markets.
- Integrated multi-source data and enabled year-over-year trend analysis with dynamic filters and KPIs, improving stakeholder insight by 30%.
- Highlighted EV leadership in countries like Norway and Iceland, showing how renewable adoption drove emission reductions of up to 40% in some markets.

U.S. Employment & Education Market Story | Story

- Built a Tableau Story analyzing education levels, unemployment, and job availability across 50 U.S. states
- Increased engagement by 40% compared to static reports by using narrative-driven visualizations and interactive filters for state-level insights.
- Identified top states (e.g., Utah, Nebraska) with <3% unemployment abd high job openings per capita, supporting career planning and workforce policy.

Airline Customer Satisfaction Prediction | (Logistic Regression & Random Forest) | Documentation

- Analyzed **129K+ airline customer records** to identify key drivers of passenger satisfaction using **EDA, Logistic Regression, and Random Forest models**.
- Built an interpretable **logistic regression model (82.8% accuracy)** to quantify the impact of loyalty status, travel class, service quality, and delays.
- Developed a **Random Forest classifier achieving 95.7% accuracy**, capturing nonlinear relationships and feature interactions for high-precision prediction.
- Performed **feature importance analysis**, revealing inflight entertainment, seat comfort, online booking, and onboard service as top satisfaction drivers.
- Delivered **actionable business insights** to prioritize service quality investments over operational delay reductions for improving customer experience.

Sales Performance Regression Analysis | (Simple & Multiple Linear Regression) | Documentation

- Conducted **regression analysis on multi-location sales data** to evaluate the impact of advertising spend, salesperson experience, and regional economics.
- Built a **simple linear regression model explaining ~99% of sales variance ($R^2 \approx 0.99$)**, identifying advertising budget as the dominant driver.

- Compared simple vs. multiple regression models using **ANOVA**, demonstrating no significant performance gain from additional predictors.
- Created **forecasting scenarios** to estimate sales outcomes under varying advertising budgets for strategic planning.
- Delivered **data-driven recommendations** enabling organizations to optimize marketing spend with high confidence and interpretability.

The 2-Degree Blueprint: Climate Policy & Emissions Forecasting | (EN-ROADS, Time Series, Regression & Optimization) | Documentation

- Designed a climate policy simulation project using EN-ROADS to evaluate global emissions pathways through 2100.
- Applied time-series forecasting (5-Year SMA & Exponential Smoothing) to project long-term greenhouse gas emissions under multiple policy scenarios.
- Performed sensitivity analysis to rank climate levers, identifying renewables, deforestation reduction, and carbon pricing as highest-impact actions.
- Built a multiple regression model ($R^2 \approx 0.82$) to quantify how policy levers jointly influence end-of-century emissions outcomes.
- Executed a regression-based optimization strategy to identify a policy configuration capable of limiting warming close to 2°C, supporting evidence-based climate decision-making.

Honors & Awards

Iron Viz Data Visualization Competition — Winner
- Central Michigan University, College of Business Administration | Jun 2025
- Recognized for excellence in data visualization, Tableau storytelling, and analytical insight at the university level.

On The Spot Award
- Tata Consultancy Services (TCS) | Feb 2024
- Awarded for immediate impact and outstanding performance in delivering high-quality analytics and data engineering solutions.

Applause Award
- Tata Consultancy Services (TCS) | Jul 2023
- Recognized for consistent high performance, collaboration, and client impact across project deliverables.

Star Team Award
- Tata Consultancy Services (TCS) | Oct 2022
- Awarded for exceptional team contribution in successfully delivering complex data initiatives.

Applause for Team Award
- Tata Consultancy Services (TCS) | Aug 2022
- Recognized for cross-functional teamwork and execution excellence.

Best Team Award
- Tata Consultancy Services (TCS) | Jun 2022
- Honored for top-performing project team, meeting aggressive timelines and quality benchmarks.

Certifications
- Tableau Desktop Specialist – Tableau
- Google Data Analytics

Education
- **Central Michigan University, Mount Pleasant, MI**
  Master of Science in Information Systems | Anticipated Graduation: Aug 2026 | GPA: 3.9
  Selected Classes: Application Development, Information Systems, Communication in Professional Contexts, Business Decision Modeling, Data Visualization: Theory and Practice, Systems Analysis and Design, Database Management for Business Systems, Enterprise Sys Mgmt/SAP Software, Business Data Analytics, Information Systems Project, Project Management Fundamentals, App of Project Manage Principles.