



RAAJVEER SUTAR

DATA ENGINEER

CONTACT

- 9518939288
- raajveersutar@gmail.com
- New Sangvi, Pune
- www.reallygreatsite.com
- github.com/Raajveers
- linkedin.com/in/raajveer-sutar

SKILLS

- Programming & Frameworks :
Python, SQL, Pandas, NumPy, PySpark,
Java, Spring Boot, Hibernate
- Data Engineering & Processing :
ETL, ELT, Data Pipelines, Data Modeling,
Data Validation, Data Governance
- Cloud Platforms :
AWS (S3, Glue, Redshift, Athena)
- Databases & Warehousing :
PostgreSQL, MySQL
- DevOps & Automation :
Docker, Jenkins, Terraform, CI/CD, Git
- API Development & Frameworks :
Django, Flask, FastAPI
- File Formats : CSV, JSON, XML

EDUCATION

- MES Abasaheb Garware Clg
MCA | 2021 - 2022 | 78%

CERTIFICATION

- Amazon web service (2021)
- Java Full stack Developer
(2023) - Symbiosis

LANGUAGES

- English
- Hindi
- Marathi



PROFILE

Data Engineer with 2.6 years of experience designing and implementing scalable data solutions across cloud environments. Proficient in Python, PySpark, AWS, SQL, and ETL pipelines, with strong expertise in data modeling, migration, and governance. Skilled at automating workflows using Docker, Jenkins, and Terraform, integrating software engineering best practices into data-driven systems. Passionate about building reliable, efficient, and secure data architectures that empower business intelligence and analytics.



WORK EXPERIENCE

Square Vision Technology

APRIL 2023 - PRESENT | 2.6 YRS

Data Engineer

- Automated Customer Identification: Built a system to automatically identify customers from logistics data files (CSV, XLSX) received via email, reducing manual effort by ~65%.
- Data Storage & Cloud Integration: Stored incoming files in AWS S3 (Landing Zone) for structured organization and cost-efficient storage.
- Orchestration Pipelines: Processed and mapped customer data, matched headers for database ingestion, ensuring accurate and reliable data flow.
- Database Management: Saved processed data in Couchbase (NoSQL) for flexible, scalable storage and fast retrieval.
- Machine Learning Integration: Predicted logistics metrics (delivery dates, CBM, cargo utilization), optimizing operations and increasing profitability by ~25%.
- ETL Automation with Jenkins & Python: Built automated ETL pipelines with version control (GitHub) and dependency management.
- CI/CD Implementation: Developed Jenkins pipelines and Dockerized ETL workflows, automating environment setup, testing, and deployment, reducing manual intervention by ~60%.
- Infrastructure as Code: Leveraged Terraform to provision cloud resources for automated pipelines, improving deployment reliability by ~50%.
- Workflow Automation: Monitored emails and triggered workflows using AWS Lambda for continuous, hands-off data processing.
- Collaboration & Governance: Partnered with cross-functional teams to implement data governance, validation, and quality frameworks, ensuring compliance and accuracy.
- Skills & Tools: Python, PySpark, SQL, Pandas, NumPy, Docker, Jenkins, Terraform, Git, AWS (S3, Lambda, Glue), Couchbase, ETL, CI/CD, Data Pipeline Automation, Machine Learning, Data Governance