

Badrish Madapuji Srinivasan

Bremen, Germany | +49 176 87452640 | badrish41@gmail.com
linkedin.com/in/badrishms | github.com/baddy1411

Target: Werkstudent Data Engineering / MLOps (16-20h/week)

PROFESSIONAL SUMMARY

Master's student in Data Engineering with 2.5 years of industry experience building **ETL pipelines, real-time data systems, and ML infrastructure** in production environments. Seeking a **Werkstudent position** to apply hands-on expertise with **Python, SQL, Apache Spark, Kafka, Airflow, and AWS**. Designed and operated pipelines processing **50K+ IoT messages/day** and **2TB+ financial data**, with a strong focus on reliability and monitoring.

TECHNICAL SKILLS

Languages: Python (Advanced), SQL (PostgreSQL, MySQL, MS SQL), Bash, DuckDB

Data Engineering: Apache Spark, Kafka, Airflow, dbt, ETL/ELT Design, Data Modeling, Real-time Streaming

Cloud & MLOps: AWS (S3, EC2, Lambda, Glue, SageMaker), Docker, Kubernetes, Terraform, CI/CD, MLflow

ML & AI: TensorFlow, PyTorch, Scikit-learn, LangGraph, Ollama, RAG Pipelines, Vector DBs (ChromaDB)

Visualization: Power BI, Tableau, Grafana, Streamlit

Spoken Languages: English (C1), German (A2 – actively learning), Tamil (Native)

EXPERIENCE

Riot Labz Pvt. Ltd.

Jan 2024 – Jul 2024

Noida, India

Software Developer (Data Infrastructure)

- Built a **FastAPI**-based ingestion service for real-time IoT telemetry from 100+ edge devices via **MQTT (50K+ messages/day)**, enabling reliable monitoring and analytics for production systems.
- Migrated legacy flat-file storage to **MySQL** with optimized indexing strategies, reducing query latency by **60%** for time-series sensor data and significantly improving dashboard responsiveness.
- Developed **Python** validation scripts to detect anomalies in sensor readings, preventing **15%** of corrupted data from entering production pipelines and reducing downstream errors.

Capgemini

Feb 2022 – Nov 2023

Bengaluru, India

Software Engineer / Analyst

- Designed and deployed **ETL pipelines** for major financial clients processing **2TB+** of transaction data using **SQL** (CTEs, window functions) and Python, ensuring reliable data delivery for regulatory reporting.
- Automated reporting workflows using **Pandas** to replace manual Excel processes, generating **Power BI** datasets automatically and reducing reporting turnaround time by **15%**.
- Collaborated with business stakeholders and engineering teams to translate functional requirements into technical specifications and data models for analytics products.

Bosch India

Oct 2020 – Jan 2021

Kanchipuram, India

Graduate Engineer Trainee

- Monitored manufacturing equipment sensors and collected production data for quality analysis, identifying recurring issues in assembly line processes.
- Automated repetitive Excel-based reporting tasks using **Python** scripts, saving **5+ hours per week**.

SELECTED DATA ENGINEERING & ML PROJECTS

Silicon Content Forecasting (with SMS group) | TensorFlow, Spark, Kubernetes

Sep 2024 – Present

- Developing **LSTM/Transformer** models to predict blast furnace silicon content on **1M+ time-series sensor records**, achieving a **12% error reduction** compared to baseline rule-based methods.
- Built scalable preprocessing pipeline using **Apache Spark** to handle noisy metallurgical sensor data.
- Deployed model evaluation framework on **Kubernetes** to benchmark multiple neural network architectures.

Predictive Maintenance Pipeline | MLflow, Airflow, AWS SageMaker

Apr 2024 – Jun 2024

- Developed an **end-to-end MLOps pipeline** for the **NASA CMAPSS dataset**, automating data preprocessing, model training, and evaluation with **Airflow** orchestration.
- Containerized training logic with **Docker** and tracked experiments using **MLflow** (15+ runs).
- Implemented **CI/CD via GitHub Actions** to deploy models to **AWS SageMaker** endpoints upon 85%+ accuracy.

Real-Time Stock Market Data Pipeline | Kafka, Spark Streaming, Grafana

Aug 2023 – Oct 2023

- Ingested live stock prices from Finnhub API using **Kafka**, processing **10K+ events per minute**.
- Processed streams with **Spark Structured Streaming** to detect price anomalies and visualized via **Grafana**.

Multi-Agent Logistics Optimizer | *LangGraph, DuckDB, Streamlit* Oct 2024 – Nov 2024
• Built a multi-agent prototype where autonomous agents query inventory tables to propose reorder schedules.

- Developed **Streamlit** interface to visualize agent reasoning chains (Chain of Thought) for debugging.

GDPR-Compliant Document QA System | *Ollama, RAG, ChromaDB* Jul 2024 – Aug 2024
• Implemented privacy-first QA system using local **Llama 3** models via **Ollama** for 500+ internal PDFs.

- Built **RAG pipeline** with **ChromaDB**, achieving relevant answers within 2–3 seconds.

EDUCATION

| | |
|---|--|
| Constructor University (formerly Jacobs University) <i>Master of Science in Data Engineering</i> Relevant Coursework: Distributed Systems, Stream Processing (Kafka), Cloud Architecture, MLOps. | Sep 2024 – Jun 2026 Bremen, Germany |
| Sri Ramakrishna Engineering College <i>Bachelor of Engineering in Mechanical Engineering</i> | Aug 2017 – May 2021 Coimbatore, India |

CERTIFICATIONS & AVAILABILITY

Certification: AWS Certified Data Engineer – Associate

Work Status: Student Residence Permit (Authorized for 20h/week). Location: Bremen (open to remote).