🔗 yadavgkp.github.io

AKASH YADAV

Data Engineer | Python Developer | Cloud & Big Data Specialist

@ akashyadav1217@ug.cusat.ac.in   📞 +49-155-10680242   ✉ Siemenstrasse 50A, Berlin
in linkedin.com/in/yadavgkp   github.com/yadavgkp

EXPERIENCE

Data Engineer – Infrastructure & Data Applications
Nocturne GmbH

📅 Nov 2023 – Sep 2025   📍 Berlin, Germany

- Built and orchestrated ETL pipelines with Python, Airflow, and Pandas — powering analytics and ML for healthcare data platforms.
- Migrated batch ETL to Databricks Delta Lake; delivered ACID guarantees and time-travel for audits.
- Designed scalable ingestion APIs with Django REST for ML model consumption and downstream systems.
- Optimized PostgreSQL queries and schema designs to reduce query time and improve reporting efficiency.
- Containerized workflows with Docker; deployed via AWS ECS, ensuring high availability and fault-tolerance.
- Automated CI/CD for pipelines and ML deployments, reducing release cycles and deployment errors.
- Collaborated with data scientists to ensure reliable feature engineering pipelines for production ML models.

Associate Software Engineer – Data & Platform
National Payments Corporation of India (NPCI)

📅 Jul 2022 – Oct 2023   📍 Chennai, India

- Developed ETL workflows to process high-volume financial transactions, supporting fraud detection and reporting systems.
- Built Spark Streaming & Kafka pipelines for anomaly monitoring in real-time payments.
- Designed and optimized PostgreSQL schemas powering analytics and machine learning features.
- Developed APIs to serve processed datasets and ML insights within payment systems.
- Enhanced data quality and observability with automated alerts, dashboards, and error handling mechanisms.
- Supported platform migration efforts, improving scalability of core payment data pipelines.
- Collaborated with cross-functional teams to align data infrastructure with compliance and regulatory standards.

Data Engineering Internships & Projects
CUSAT

📅 Jul 2019 – Jun 2022   📍 Kerala, India

- Implemented ETL workflows with Python, Pandas, and SQL for academic and small-scale industry projects.
- Built applications such as an Electricity Billing System (Django + MySQL) and CCTV Activity Summarizer (OpenCV).
- Developed personal machine learning projects including predictive models and classification pipelines using Python, scikit-learn, and Pandas.

TECHNICAL SKILLS

Programming & Frameworks

- Python (Pandas, NumPy, FastAPI, Django)
- PySpark & Spark Streaming
- SQL & PostgreSQL Optimization
- REST API Development
- Jupyter, VS Code, PyCharm
- Data Modeling & Schema Design

Data Engineering & Cloud

- ETL Pipeline Development
- Apache Airflow & Workflow Orchestration
- Databricks & Delta Lake
- AWS (EC2, RDS, S3, ECS, Lambda)
- BigQuery, Redshift
- Real-Time Streaming (Kafka, Spark Streaming)

DevOps & Infrastructure

- Docker & Containerization
- Kubernetes Basics
- Terraform (IaC)
- Infrastructure Optimization & Scaling

Tools & Practices

- Git & Version Control
- Agile/Scrum & Jira
- Linux/Unix Systems
- Postman & API Testing
- Data Documentation & Governance

KEY PROJECTS

- **PDF Q&A App with LLMs** – RAG pipeline using LangChain, GPT-3.5, FastAPI, Streamlit.
- **Medical AI Platform** – MLOps pipeline (Docker, MLflow, AWS) for model deployment.
- **Real-time UPI Fraud Detection** – Spark Streaming + Kafka pipeline for anomaly detection in transactions.
- **CCTV Activity Summarizer** – Computer Vision (YOLO, OpenCV) for summarizing video events.
- **Electricity Billing System** – Django + MySQL-based ETL and billing solution.

EDUCATION

PG Program in AI and Machine Learning
IIIT Hyderabad

📅 Mar – Nov 2023   📍 Hyderabad, India

- Applied Python for data engineering and machine learning implementations.
- Built projects using Pandas, NumPy, and end-to-end ML/data workflows.

- -

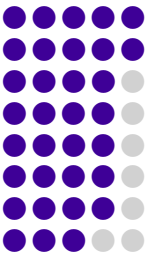### B.Tech in Computer Science and Engineering
Cochin University of Science And Technology (CUSAT)

📅 Jul 2018 – Jun 2022   📍 Kerala, India

- GPA: 7.63/10
- Coursework included Database Systems, Big Data, Distributed Systems.

TECHNICAL PROFICIENCY

Python, Pandas, NumPy   ●●●●●
SQL & PostgreSQL Tuning   ●●●●●
Apache Airflow   ●●●●○
Spark Streaming & Kafka   ●●●●○
ETL & Data Modeling   ●●●●○
Docker & AWS (EC2, RDS, ECS, S3) ●●●●○
REST API Development   ●●●●○
CI/CD Pipelines   ●●●○○

AWARDS & RECOGNITION

- **Excellence in Ownership Award** – Nocturne GmbH (for infrastructure/data projects).
- **Star Performer of the Month** – National Payments Corporation of India (NPCI).
- First Prize (Team) – NSS Community Development Program.
- Winner (Doubles) – Intra-University Badminton Championship, CUSAT.
- Best Speaker Award – Intra-College Debate Competition, CUSAT.

KEY STRENGTHS

ETL Design & Optimization
Real-Time Data Pipelines
System Health Monitoring
Performance Optimization
Cross-functional Collaboration
Problem-Solving   Attention to Detail

LANGUAGES

- **English:** Fluent
- **German:** Beginner (actively learning - A1)