

AKASH YADAV

Databricks & Python Data Engineer | Azure Data & Cloud Specialist

@ akashyadav1217@ug.cusat.ac.in
in linkedin.com/in/yadavgkp

+49-155-10680242
github.com/yadavgkp

Siemenstrasse 50A, Berlin, Germany



EXPERIENCE

Databricks Data Engineer – Cloud Data Infrastructure Nocturne GmbH

Nov 2023 – Apr 2025

Berlin, Germany

- Led design and implementation of a **data lakehouse platform** using **Azure Databricks**, **Delta Lake**, and **PySpark**.
- Built scalable **ETL and ELT pipelines** with **Azure Data Factory**, **Python**, and **SQL** for analytics and ML workloads.
- Integrated security and access control through **Azure Key Vault** and role-based configurations.
- Migrated legacy batch workflows to modern **Databricks Delta architecture**, enabling ACID compliance and time-travel.
- Automated data ingestion, validation, and transformation using **Airflow** and **Databricks Jobs**.
- Deployed containerized data services with **Docker** and **Azure DevOps CI/CD pipelines**.
- Collaborated with data scientists to productionize ML models and optimize feature pipelines.
- Enhanced data reliability, monitoring, and audit compliance across the data lifecycle.

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

Jul 2022 – Oct 2023

Chennai, India

- Developed and maintained **PySpark**-based ETL pipelines for processing high-volume financial data streams.
- Engineered **Kafka** and **Airflow** pipelines for near real-time fraud detection and anomaly monitoring.
- Optimized **PostgreSQL** and **Azure SQL** schemas, improving reporting and query performance by 30%.
- Integrated data products into the Azure ecosystem, including **Data Factory**, **Blob Storage**, and **Key Vault**.
- Contributed to modernization projects migrating on-prem workloads to **Azure Databricks** and **Delta Lake**.
- Implemented data validation, monitoring dashboards, and alerting for production reliability.
- Collaborated with DevOps teams to ensure scalable and secure deployments across environments.

Data Engineering Internships & Academic Projects Cochin University of Science and Technology (CUSAT)

Jul 2019 – Jun 2022

Kerala, India

- Designed end-to-end ETL pipelines using **Python**, **Pandas**, and **SQL** for academic research datasets.
- Developed web applications such as an **Electricity Billing System (Django + MySQL)** and a **CCTV Activity Summarizer (OpenCV)**.
- Created **machine learning models** for predictive analytics using **scikit-learn** and **TensorFlow**.

TECHNICAL SKILLS

Programming & Frameworks

Python (Pandas, NumPy, FastAPI, Django)

PySpark & Spark Streaming

SQL & PostgreSQL Optimization

REST API Development

Data Modeling & Schema Design

Data Engineering & Cloud

Databricks & Delta Lake

Azure Data Factory & Key Vault

Azure Synapse / Blob Storage

ETL Pipeline Development

Apache Airflow & Workflow Orchestration

AWS (EC2, RDS, S3, ECS, Lambda)

DevOps & Infrastructure

Docker & Containerization

Terraform (IaC)

CI/CD (Azure DevOps, GitHub Actions)

Cloud Security & Access Management

Tools & Practices

Git & Version Control

Agile/Scrum & Jira

Linux/Unix Systems

Postman & API Testing

Data Documentation & Governance

Testing & Debugging (Pytest, Unittest)

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

IIIT Hyderabad

📅 Mar – Nov 2023 📍 Hyderabad, India

📍 Hyderabad, India

- Applied **Python**, **NumPy**, and **Pandas** for ML-based data engineering.
- Built end-to-end ML pipelines and deployed models on cloud environments.

Cochin University of Science and Technology (CUSAT)

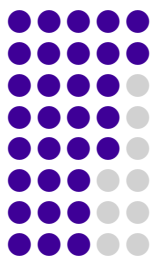
Jul 2018 – Jun 2022 Kerala, India

📍 Kerala, India

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

TECHNICAL PROFICIENCY

- Python, PySpark, Databricks, SQL
- Azure (Data Factory, Key Vault, Delta Lake)
- Airflow, Kafka, ETL & Data Lakehouse Design
- PostgreSQL & Azure SQL Optimization
- Docker, Azure DevOps, CI/CD Pipelines
- Machine Learning Pipelines & Integration
- REST API Development & Django
- Monitoring, Logging & Automation



AWARDS & RECOGNITION

- **Excellence in Ownership Award** – Nocturne GmbH (for leadership in data infrastructure initiatives).
- **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)