



Arpit Angraj

Data Engineer

6+ Years of Experience

DE Permanent Residence

(Niederlassungserlaubnis)

Marktoberdorf, Bavaria, Germany

+49-1727484721

arpitangraj@gmail.com

linkedin.com/in/arpitangraj

PROFESSIONAL SUMMARY

Databricks-certified Data Engineer with 6+ years of experience across consulting, manufacturing and fintech, building data systems used for day-to-day operations, reporting, and financial decision-making. Comfortable moving between hands-on engineering, architecture, and conversations with decision-makers. Curious by nature, quick to pick up new technologies, and happiest when that learning turns into something useful.

CORE SKILLS

Databricks & Spark | Python, PySpark & SQL | AWS Data Engineering | Streaming, Kafka & Change Data Capture (CDC) | Client Delivery & Production Support

WORK EXPERIENCE

Payla Services GmbH – Munich, Germany

Data Engineer

03/2025 – Present

- **Owned and operated the data platform as a solo Data Engineer** from June 2025, partnering directly with C-level stakeholders and collaborating with Finance, Risk, Integration, and SRE teams on production reliability, reporting, migrations, and incident resolution.
- **Engineered PySpark and Polars pipelines** across the medallion architecture for batch, streaming, Change Data Capture (CDC), reconciliation, and near-real-time financial data processing.
- **Built a GitHub Copilot custom agent** to identify Databricks jobs with high memory usage and propose Kubernetes resource changes, changelog updates, and validation steps, then showcased the approach to other engineering teams for day-to-day platform use.
- **Collaborated with SRE and engineering teams** to resolve production incidents across Amazon MSK, Kafka Connect, Databricks Structured Streaming, MongoDB CDC, and Kubernetes during migrations and critical reporting cycles.
- **Optimized Kubernetes resource allocation** for 50+ bronze tables, configuring CPU utilization and memory and reducing infrastructure costs.
- **Developed Dagster-based financial reporting pipelines** for backup servicing partners, delivering one-time and daily bank reporting files to AWS S3 through Dagster and Databricks.
- **Automated deployments using GitLab Continuous Integration and Continuous Delivery (CI/CD)** and merge trains, publishing Python wheels to AWS S3 and managing versioned job releases across AWS and Kubernetes.
- **Prepared 100+ unit and integration tests** with Pytest, LocalStack, Faker, and schema validation, achieving approximately 95% coverage across medallion-layer transformations, Dagster orchestration, and Databricks SDK integrations.

- **Built AGCO’s Enterprise Data Hub on Databricks from scratch**, helping migrate workloads from AWS-native data services to a Databricks-based platform on AWS and establishing Unity Catalog, access management, Databricks Jobs, Git integration, Databricks Asset Bundles (DABs), and GitHub Actions.
- **Migrated and optimized Spark, SQL, Kafka, and Databricks workloads**, improving reporting speed by 30%, reducing latency by 25%, and improving Unity Catalog query performance by up to 3×.
- **Proposed and delivered an AWS AppFlow-based Salesforce ingestion architecture**, replacing long-running custom REST API jobs in Glue and reducing Salesforce ingestion costs by more than 90%.
- **Ran Databricks workshops and training sessions for data engineers and business teams**, helping teams understand and adopt the new Enterprise Data Hub.

Infosys Ltd. – Munich, Germany

Data Engineering Consultant | Client: AGCO GmbH / Fendt, Marktoberdorf

Technology Analyst

Senior Systems Engineer

Systems Engineer

01/2023 – 03/2024

01/2022 – 12/2022

12/2019 – 12/2021

- **Worked as a billable Data Engineering Consultant for AGCO/Fendt across multiple teams**, delivering cloud and data solutions directly with client engineering and business teams.
- **Delivered work in Scrum teams**, translating requirements into user stories, estimating effort with story points, and managing delivery across sprints.
- **Optimized Spark and SQL workloads** through caching, partitioning, and workflow tuning, improving processing performance by 20%.

TECHNOLOGY STACK

Databricks & Distributed Data: Databricks, Apache Spark, PySpark, Spark SQL, Spark Structured Streaming, Delta Lake, Unity Catalog, Databricks Asset Bundles, Polars

Programming, Testing & Delivery: Python, SQL, REST APIs, Pytest, GitLab CI/CD, GitHub Actions, Terraform, Docker, Kubernetes

Cloud, Databases & Orchestration: Amazon Web Services (AWS), AWS S3, AWS Glue, AWS Lambda, AWS Redshift, PostgreSQL, MongoDB, Apache Kafka, Kafka Connect, Dagster

Architecture & Data Engineering: ETL, ELT, Batch Processing, Stream Processing, CDC, Medallion Architecture, Data Warehousing, Data Modeling, Data Governance, Data Quality, Data Reconciliation

Applied AI: Databricks Genie, GitHub Copilot, AI Agent Development

Domain: Fintech & Payments, Manufacturing, Industry 4.0, Industrial IoT

EDUCATION & CREDENTIALS

Bachelor of Engineering — Information Science
Visvesvaraya Technological University, Bengaluru, India

07/2015 – 07/2019

Certification: Databricks Certified Data Engineer Associate

Languages: English — Fluent | German — A1, currently learning B1