

# SZYMON KOSTER

Data & Platform Engineer

## CONTACT

- +48 698 113 080
- szymonkoster@gmail.com
- linkedin.com/in/szymonkoster/

## ABOUT ME

Data & Platform Engineer with deep expertise in the Azure Databricks ecosystem. Proven specialist in modernizing data platforms (Lakehouse, Unity Catalog) and building scalable Internal Developer Platforms (IDP). Expert in bridging the gap between Data Engineering and MLOps to deliver measurable business value.

## CAREER HIGHLIGHTS

- Optimized platform costs by over 60% by architecting and executing a full-scale migration of a legacy Data Hub (SQL Server) to an Azure Databricks Lakehouse.
- Scaled data governance for 300+ users by implementing Unity Catalog from scratch, enforcing fine-grained RLS, and ensuring Antitrust compliance across multi-market environments.
- Achieved near-real-time data ingestion by designing end-to-end ETL pipelines leveraging Change Data Capture (CDC), PySpark Structured Streaming, and Auto Loader.

## TECHNICAL SKILLS

- Core & Data Processing:** Python (Advanced), SQL (Spark SQL, T-SQL), PySpark, Pandas, Bash, Delta Live Tables
- Platform & Cloud:** Databricks, Unity Catalog, Delta Lake, Azure (ADF, ADLS Gen2, Machine Learning), GCP (Vertex AI)
- DevOps & MLOps:** Terraform (IaC), Docker, MLflow, GitHub Actions, Azure DevOps, Databricks Asset Bundles (DABs)
- Code Quality & Security:** Ruff, Mypy, SonarQube, Snyk, OAuth2
- Project Management:** Scrum, Jira, Confluence, Azure Boards, Trello
- Languages:** Polish (native), English (full business proficiency)

## PROFESSIONAL EXPERIENCE

### Data & ML Platform Engineer | Cloudfide [04.2025 – present]

**Enhancing Internal Developer Platform for GenAI/MLOps.** Engineered the Azure Databricks integration for an enterprise ML accelerator platform (IDP), enabling teams to develop and execute platform-agnostic pipelines across AzureML, Databricks, and Google Vertex AI. Standardized deployment strategies by migrating legacy pipelines to Databricks Asset Bundles (DABs) (YAML-based IaC), establishing a new standard for daily automated releases. Resolved critical cross-cloud authentication bottlenecks by implementing robust OAuth2 patterns between Databricks, Azure, and internal services. Maintained high code quality standards by enforcing CI/CD gates with Ruff, SonarQube, and Snyk security scans.

### Mid Data Engineer | Cloudfide [09.2022 – 04.2025]

**Enterprise Data Lakehouse Migration & Modernization.** Architected and executed a full-scale migration of a legacy Data Hub (SQL Server/AAS) to Azure Databricks Lakehouse, handling 50GB+ semantic models. Designed end-to-end ETL pipelines leveraging Change Data Capture (CDC) to trigger PySpark Structured Streaming and Auto Loader jobs, enabling near-real-time data ingestion and significantly reducing latency for business reporting in PowerBI. Secured data governance for 300+ users by implementing Unity Catalog from scratch, enforcing RLS and Antitrust compliance. Optimized platform costs by introducing FinOps practices, including cluster rightsizing, adoption of Serverless SQL warehouses, and automated resource termination policies.

### Junior DevOps/DataOps Engineer | Cloudfide [03.2022 – 09.2022]

Maintained high availability of critical data platforms, reducing ServiceNow incident resolution time by implementing proactive monitoring dashboards (Databricks Ganglia, Azure Monitor) with automated email alerting. Collaborated with global teams in a 100% remote environment to resolve complex L3 production incidents affecting downstream business reporting.

## EDUCATION

AGH University of Krakow

### Master of Science in Applied Computer Science [2024 – 2026]

Thesis: "Optimization and hypertuning of supervised and unsupervised models in the context of Data Engineering."

### Bachelor of Engineering in Computer Science [2020 – 2024]

Thesis: "A cross-platform mobile application based on GenAI technology to generate character cards for people involved in historical reenactment."

## KEY CERTIFICATIONS & COURSES

- Databricks Certified: Data Engineer Professional
- Databricks Certified: Data Engineer Associate
- Databricks Certified: AI Agent Fundamentals
- Databricks Certified: Lakehouse Fundamentals
- Microsoft Certified: Azure Fundamentals (AZ-900)
- Umiejętności Jutra (Google & SGH): Focused on building and deploying AI Agents using Google Cloud ecosystem.



## ADDITIONAL ACTIVITIES

**AI Automation Enthusiast:** Actively developing personal productivity workflows using Activepieces and LLM-based agents, applying advanced prompt engineering techniques.

**Medieval Reenactor:** Participating in historical reenactments, enhancing my analytical thinking and conscientiousness, combined with research skills and project management.