

RAKSHAN BHANDARY

Data Engineer | Python, PySpark, SQL, Azure, Databricks, Kafka, PostgreSQL

Email: bhandaryrakshan@gmail.com

Phone: +49 1525 8427188

LinkedIn: linkedin.com/in/rakshan-bhandary-589380187

Website: rakshanbhandary.pages.dev

Location: Freiburg, Germany (open to relocation)

PROFILE

Data Engineer with 5 years of experience (3 years full-time at Accenture, 2 years as working student in Germany) building batch and streaming data pipelines in production. Reduced end-to-end processing latency from 30 minutes to about 10 seconds by helping redesign a batch system as a streaming architecture on Azure. Sole architect of an automated software compliance platform at SensoPart that cut recurring effort per release from 2 weeks to a few days, serving the software supply chain of major automotive manufacturers. Completing an M.Sc. in Embedded Systems Engineering at the University of Freiburg; available full-time from October 2026.

CORE SKILLS

- **Programming:** Python (pandas, matplotlib), PySpark, SQL, Streamlit, Linux/Unix shell scripting, YAML
- **Azure:** Data Factory, Synapse Analytics, Event Hub, ADLS Gen2, Functions, Key Vault, Blob Storage
- **Data Engineering:** ETL/ELT pipelines, streaming and event-driven architectures (Kafka), batch processing, API-based ingestion, reporting automation
- **Data Modeling:** dimensional modeling (Star Schema, Snowflake Schema, OBT), schema design
- **Databases:** PostgreSQL, Oracle, MongoDB
- **Platforms & DevOps:** Databricks, Docker, GitLab CI/CD, Azure DevOps, Git, Informatica PowerCenter
- **Currently learning:** dbt, Snowflake

PROFESSIONAL EXPERIENCE

Data Engineer (Working Student) | SensoPart GmbH, Germany

Mar 2024 – Jun 2026

Sole architect and engineer of the company's automated open-source compliance and Cyber Resilience Act (CRA) platform, from design through production deployment. Serves the software supply chain of major automotive manufacturers.

- Replaced a fully manual, UI-based license scanning workflow with an automated end-to-end pipeline built on FOSSology, reducing recurring compliance effort from about 2 weeks to a few days per release.
- Designed an incremental scan reuse architecture: segmented 2-3 GB release codebases into 10-15 components so previously scanned and curated results are automatically detected and reused, eliminating redundant re-scanning across release cycles.
- Built a PostgreSQL-based single source of truth for curated license and copyright data, generating all client report formats (SPDX, Excel, plain text, client-specific formats) dynamically from one dataset.
- Built a Streamlit web application so colleagues could review and edit the platform's PostgreSQL data through a browser without direct database access, with role-based permissions (streamlit-authenticator) and rules restricting which operations each access level could perform.
- Built ingestion pipelines for CVE vulnerability data from the NVD API and Yocto build manifests, combining both sources for complete vulnerability coverage and automated SBOM generation.
- Integrated the pipeline into GitLab CI/CD with the DevOps team, triggering scans and report generation automatically on each production release.

Data Engineer / Software Engineer | Accenture, India

Oct 2020 – Sep 2023

Worked across a legacy ETL estate and a greenfield Azure streaming platform for a large enterprise client, from proof of concept to production.

Modern Azure Data Platform

- Designed and developed PySpark streaming pipelines ingesting, transforming, and standardizing JSON data from multiple heterogeneous sources using Azure Event Hub, Kafka, and ADLS Gen2.
- Reduced end-to-end data processing latency from 20-30 minutes (batch) to about 10 seconds by helping design a hybrid architecture combining near-real-time streaming with batch retry logic for failed events.
- Contributed to implementing a standardized JSON schema (design led by the team) that normalized multi-source data into one unified structure, allowing all sources to flow through the same pipeline.
- Implemented config-driven Spark SQL transformations, enabling new queries and logic changes without code deployments.
- Enriched streaming order events in real time by joining against PostgreSQL-hosted customer master data, delivering complete records downstream without additional lookups.
- Led the integration of independently developed pipeline modules into one end-to-end system, coordinating code changes across teams.
- Orchestrated batch and streaming workloads with Azure Data Factory (scheduling, dependency resolution, automated triggers).

Legacy ETL Systems

- Maintained Informatica-based ETL workflows and Oracle batch processing systems (commission calculations, roughly 30-minute cycles).
- Developed Unix shell scripts for data validation, checksum verification, and date integrity checks on incoming files before ingestion.
- Extended existing ETL workflows to onboard new data sources, adding field-level transformations and loading logic for new file and table structures.
- Provided production support, troubleshooting and resolving pipeline failures in coordination with the operations team.

EDUCATION

M.Sc. Embedded Systems Engineering | University of Freiburg, Germany Oct 2023 – Aug 2026 (expected)

GPA 1.8 (German scale).

B.E. Electronics and Communication Engineering | NMAM Institute of Technology, India 2016 – 2020

CGPA 9.01/10.

LANGUAGES

English — C1 (professional working language)

German — A1 (actively learning)

Other — Hindi, Kannada, Tulu (native/fluent)