Curriculum vitae

M.Sc. Soheil Nezakat

Personal information

Date of birth: 22. February 1990

Nationality: Iranian

Address: Rollstr 5
38678 Clausthal-Zellerfeld

Phone: 0179 / 2023476

Email: contact@soheiln.tech

Profile

I am a research-oriented data engineer with a strong focus on relational database systems, data profiling, and metadata discovery. I have hands-on experience working with large-scale relational data, including schema design, indexing, CDC-based ingestion, and scalable data pipelines using Apache Kafka, Apache NiFi, and clustered database architecture.
In my research at TU Clausthal, I focus on improving data quality and structural consistency across relational and graph-based systems through metadata extraction, data validation, and dependency analysis. My master's thesis investigates database-backed structured data for reliable data-driven systems, with an emphasis on metadata discovery, consistency checking, and validation.

Research Interests

Data profiling and automated metadata discovery in relational databases; functional dependency, candidate key, and inclusion dependency analysis; database systems and schema design; algorithmic methods for exploring large search spaces in metadata discovery; knowledge graphs and graph databases (Neo4j); Java- and Python-based research prototypes for data analysis; scalable data engineering systems; data validation, consistency checking, and the reliability of data-driven systems.

Professional Career

Research Associate, Clausthal-Zellerfeld, DE — 06.2023 - ongoing
Technical University of Clausthal

- *Knowledge-Graph-Based Metadata Grounding for Reliable Data-Driven Systems (Current Thesis):*
  Current research focuses on improving the reliability and consistency of data-driven systems by grounding learned models in database-backed structured data and knowledge graphs. I design and implement Java- and Python-based backend components for structured data retrieval, graph access, and metadata management, enabling validation and consistency checks between relational data, graph representations, and downstream model outputs. The work emphasizes

metadata discovery, structural consistency, and the validation of structured information as a foundation for reliable system behavior.

- *SkyCam Data System and Cloud Behavior Prediction:*
  Worked on the SkyCam project, including data collection from sky-facing cameras, image storage, and preprocessing. Developed deep learning models to detect clouds and predict cloud movement, supported by structured data storage and temporal evaluation.

- *Data Processing Pipelines for Sensor and IoT Data:*
  Built data processing workflows in Java and Python to ingest, clean, transform, and store large volumes of sensor data. Developed backend components for data ingestion, scheduling, and database integration, with a focus on building reliable and maintainable data pipelines.

- *IoT Sensor Data Collection and Time-Series Data Management:*
  Collected data from sensors and IoT devices, including device programming, data acquisition, and reliable data transfer. Designed pipelines to store time-series data in InfluxDB, manage missing or noisy data, and support efficient querying for analysis.

- *Time-Series Monitoring and Visualization:*
  Built monitoring and visualization dashboards using Grafana on top of InfluxDB to analyze sensor data over time. Focused on data quality checks, data aggregation, and clear visualization to support research experiments and system monitoring.

Head of Software & Data Analytics, Dubai, UAE 03.2021 - 09.2022
Iranian hospital

- *Data Integration & System Optimization:*
  Designed and operated large-scale data integration pipelines using Apache NiFi and Kafka, including CDC-based ingestion, streaming workflows, and reliable data exchange between transactional and analytical systems, reducing end-to-end processing time by approximately 40%. Led an Oracle-to-PostgreSQL migration, including schema redesign, indexing strategies, and query optimization, resulting in a ~50% improvement in query performance and system reliability. Designed and maintained clustered relational database architectures, ensuring high availability, fault tolerance, and consistent data access across Oracle and MySQL systems.

- *Database Administration & Data Engineering:*
  Managed mission-critical relational databases as a DBA, covering schema evolution, index tuning, CDC pipelines, clustered deployments, backup strategies, and performance monitoring. Regularly analyzed database structures and workloads to identify redundancies and bottlenecks, contributing to measurable improvements in system stability and maintainability.

- *Software Quality & Automation:*
  Developed Python-based test and validation frameworks, achieving approximately 95% code coverage, reducing production issues by around 30%, and improving release stability for data-intensive backend components.

- *Business Intelligence & Data Visualization:*
  Built Qlik Sense and Oracle BI dashboards with key KPIs, improving decision-making by approximately 30%. Utilized Qlik Sense, Power BI, and Oracle Data Visualization to identify anomalies and clusters, enhancing analytical insights by around 20%.

- *AI & Knowledge-Based Systems in Healthcare:*
  Developed expert systems for medical diagnosis, improving efficiency by approximately 30%. Applied fuzzy logic to enhance patient data classification and optimize hospital resource utilization. Implemented knowledge-based systems to detect insurance fraud, reducing fraudulent activity by around 20%.

IT manager, Dubai, UAE 02.2020 - 02.2021
Iranian hospital

- *System Integration & Compliance:*
  Spearheaded the integration of the DHA Hasana system with API-based workflows, reducing data errors by approximately 30% and ensuring compliance. Integrated the hospital information system (HIS) with financial applications via custom middleware, improving financial accuracy by around 35% and streamlining audit processes.

- *Agile Development & Real-Time Analytics:*
  Collaborated with cross-functional teams in Agile environments, leading sprint planning, retrospectives, and the delivery of real-time analytics features.

- *Healthcare Systems & Revenue Management:*
  Implemented revenue cycle management (RCM) software, boosting revenue cycle efficiency and reducing billing errors by approximately 20%. Contributed to achieving EMRAM Stage 6 at HIMSS by deploying advanced health information systems that enhanced service quality and operational efficiency.

- *Digital Transformation:*
- Directed the hospital's transition to paperless workflows through digitization and cloud storage, reducing paper usage by approximately 90% and modernizing operational processes.

Software engineer, Dubai, UAE 12.2014 - 01.2020
Iranian hospital

- *Patient Engagement & Healthcare Systems:*
  Developed data-driven backend services integrating transactional systems, ensuring data quality, consistency, and reliable information flow. Deployed a hospital information system (HIS) with automation tools (Selenium, APIs), reducing task time and generating approximately $20 million in additional revenue in the first year. Reduced outpatient waiting times by 80% by integrating EID smart cards with the HIS using HL7 standards. Connected medical equipment with the HIS via HL7, improving data accuracy by 30% and streamlining clinical workflows.

- *Operational Efficiency & Workflow Optimization:*
  Optimized hospital operations across medical services, human resources (HR), doctors' workbench, nursing, pharmacy, CRM, and inventory systems, increasing overall efficiency by approximately 25%. Standardized HR workflows for personnel tracking and attendance, improving HR operations by around 20%. Designed and implemented data management workflows using SharePoint, InfoPath, and SQL, increasing data handling efficiency by approximately 25%.

- *Software Quality & Secure Identification:*
  Applied the V-model to ensure robust software architecture and testing in embedded systems. Implemented a first-time registration system using vein pattern recognition, improving security and identification efficiency by approximately 40%.

- *Data Engineering & Project Delivery:*
  Developed and maintained ETL workflows, performing advanced indexing, join optimization, and performance tuning on large relational datasets, resulting in approximately 30% faster performance. Delivered cross-functional projects using Agile methodologies, applying work breakdown structures and risk mitigation strategies to ensure on-time, on-budget delivery.

Education

Technical University of Clausthal, Clausthal-Zellerfeld — 10.2022 - 03.2026
M.Sc. Computer Science (*Reliable Data-Driven Systems*)

Islamic Azad University, Dubai — 03.2013 - 09.2015
M.Sc. Software Engineering (Data embedding and image processing)

Islamic Azad University, Lahijan — 09.2008 - 02.2013
B.Sc. Computer Science

Technical Skills

Programming & Research Prototyping:
Python, Java, R, MATLAB, Bash/Shell scripting; development of research prototypes and backend components for data processing and analysis

Database Systems & Data Management:
Relational database systems (PostgreSQL, Oracle, MySQL, SQL Server); schema design, indexing, query optimization, CDC-based ingestion; clustered and redundant database architectures; graph databases (Neo4j); SQL

Data Engineering & Metadata-Aware Pipelines:
Apache Kafka, Apache NiFi, ETL workflows, Spark SQL, Hadoop, Delta Lake; structured data processing, validation, and integration; Java-based data pipelines; large-scale data ingestion and transformation

Systems, Tooling & Engineering Practices:
Docker, Git, CI/CD, Linux (Ubuntu, Oracle, CentOS); REST APIs; monitoring and system observability (Grafana, PRTG); cloud platforms (AWS, Azure, OCI)

Languages

Persian:
C2
English:
C1 (IELTS Academic 6.5)
German:
B1 (currently improving taking courses)

Clausthal-Zellerfeld, 28 January 2026

Soheil Nezakat