

# Dmitrii Malashikhin

New York, NY · dmitrii@sent.com · 646.671.4652 · linkedin.com/in/themdq · github.com/themdq  
Open to Relocation · Fully Authorized to Work · No Visa Sponsorship Required

## SUMMARY

Data Engineer with 3+ years of experience building production data pipelines and ELT/ETL workflows in cybersecurity. Specialized in high-volume data ingestion, schema design, and pipeline optimization across 30+ vendor feeds and millions of records. Reduced data onboarding time by 75%, improved test coverage to 96%, and lowered production incidents through schema validation, CI/CD gates and automated checks.

## PROFESSIONAL EXPERIENCE

### Data Engineer

10/2023 – 01/2026

**R-Vision** – Cybersecurity vendor, 100+ clients, 15M+ assets

- Built and operated 30+ production data pipelines ingesting and normalizing vendor data (Microsoft, Adobe, Ubuntu), supporting millions of vulnerability records used by analytics and scanner systems.
- Enhanced new-source onboarding with base classes, Pydantic validation models and source-specific configuration, cutting development time from 1 day to 3 hours (75%) and lowering production incidents by 23%.
- Introduced test-driven development with pytest and CI/CD gates, increasing unit and integration test coverage from 70% to 96%.
- Optimized large-source rebuilds and enrichment with Python content hashing, cutting rebuild time from a few hours to approximately 5 minutes.
- Developed FastAPI-based data platform with JSON Patch, enabling bulk edits across 1M+ records and reducing manual workload by 40%.

### Data Engineer

08/2022 – 10/2023

**Datagile** – Cybersecurity vendor, 250+ clients, 600+ projects/year

- Implemented and maintained 10+ production integrations for enterprise clients, supporting \$250K+ in annual revenue by enabling reliable ingestion and normalization of heterogeneous event data.
- Designed a template-driven rule authoring framework that accelerated normalization development time by 2x and improved consistency across integrations.
- Replaced a 6-step manual release process with GitLab CI/CD pipelines, automated tests, and rollback procedures, decreasing production incidents by 65% and documentation-related support requests by 33%.
- Created packaging and release utilities adopted by 50+ engineers, saving approximately 12.5 engineering hours per day across repetitive deployment workflows.

## SKILLS

**Languages:** Python, SQL, Bash

**Data Engineering:** Databricks, dbt, Kafka, Spark, Airflow, Prefect, Data Modeling, Data Orchestration

**Databases:** PostgreSQL, Snowflake, Elasticsearch, MongoDB, Redis

**Infrastructure:** AWS (S3, RDS, Lambda), Docker, Kubernetes, GitHub Actions, GitLab CI/CD

**Observability & Reporting:** Power BI, Grafana, Prometheus, Loki, ELK Stack

**Backend & Testing:** FastAPI, Django, Flask, SQLAlchemy, Pydantic, REST APIs, asyncio, pytest

## PROJECTS

**Vulnerability Intelligence Pipeline** | Python, Prefect, dbt, Snowflake, FastAPI, Streamlit, AWS RDS

- Built an ELT pipeline for CVE ingestion using Python, Prefect, retry/backoff logic and idempotent upserts to keep repeated loads consistent and recoverable.
- Modeled vulnerability data in Snowflake with dbt staging and marts layers, creating analytics-ready datasets for severity, vendor, product and exploitability analysis.
- Shipped a Streamlit dashboard for monitoring data freshness, pipeline runs and vulnerability trends.

## EDUCATION

**Master of Science in Information Systems**, Admiral Makarov State University – 2024

**Bachelor of Science in Computer Science**, Admiral Makarov State University – 2022