András Novoszáth

Budapest, Hungary | +36 30 889 4244 | nocibambi@gmail.com | LinkedIn

Delivers **Python analytics** with researchers and engineers, **replacing legacy systems with cloud-native APIs** and resilient automation. Writes clean **pandas/numpy** code, integrates **REST workflows**, and hardens **AWS-based CI/CD** Operates in **Agile** teams, models in **SQL Server**, and keeps **monitoring, logging, and performance tuning** in focus for smooth upgrades and support.

CAREER HIGHLIGHTS

Data Platform Engineering

- Developed ETL pipelines to provide on-chain and off-chain data about **Web3, Health, and Governance**.
- Engineered **serverless microservices**, infrastructure, and ETL pipelines enhancing data flow and access with Python and AWS. Reduced workflow inefficiencies in CI/CD pipelines.
- Built out, improved, and maintained **data operation pipelines and processes**.
- Architected **monitoring and alerting** for data collection and CI/CD.

Data Science and Machine Learning

- Facilitated data-driven decision-making across **blockchain, governance, and health** sectors.
- Created customized **reporting** solutions. Bolstered funding for medical device development by building reports from clinical trials.
- Used **generative AI** to create **infrastructure and data flows documenation** for architecture review.
- Built a glucose **forecasting model** achieving the accuracy of frontline medical devices.

Web3 Data Engineering

- Developed high-impact **on-chain and off-chain analytics tools**. Built data pipelines for DAOs providing community, governance, and market insights.
- Built a point tracker evaluating **Aave** liquidity providers.
- Created a **community** to learn about and build on Web3 data.

SKILLS

- **Data Engineering**: Data Collection (APIs, Beautifulsoup, Selenium, Playwright) | Data Validation (Pydantic) | Data Processing (pandas) | Databases (Microsoft Server SQL, InfluxDB) | Flat files (csv, JSON, parquet, feather) Data Monitoring (AWS Cloudwatch) | Data Pipeline Orchestration (AWS EventBridge, Apache Airflow, AWS Step Functions)
- **Software Engineering**: Cloud Technologies (AWS) | AWS Microservices (Lambda, EC2, SNS, SQS) | Testing (pytest) | Version Control (git, Github, Gitlab) | Frontend (HTML, CSS, Javascript, Anvil) | Backend (Django, FastAPI) | Static-site generators (Jekyl, Hugo, MkDocs, Sphinx) | Python tooling (pip, conda, poetry, venv, black, flake8, mypy)
- **DevOps**: CI/CD (AWS Codebuild/Codepipeline, Github Actions) | Infrastructure as Code (Python CDK/Terraform) | Software Deployment (Docker, AWS CodeArtifact, AWS ECR) | SysAdmin (Linux, bash)
- **Data Science**: Data processing (pandas, numpy) | Data Visualization (matplotlib, seaborn, altair, bokeh, plotly) | Querying (MS SQL, BigQuery, InfluxDB, Snowflake SQL) | Time-Series Analytics (pandas, InfluxDB) | Dashboards (Streamlit, Anvil/Dash)
- **Machine Learning**: Libraries (scikit-learn, keras), Applications (prediction, clustering, forecasting, anomaly detection), Methods (multi-label classification, rebalancing, cross-validation, evaluation, feature engineering)
- **AI Engineering**: Integrated IDE workflows (VSCode, Cursor), APIs, prompt engineering (specs-driven coding, chain of thought). Developed semantic event logger.
- **Blockchain Analytics**: On-chain Analytics (Flipside, Dune, web3.py, Etherscan) | Off-chain Analytics (Discord, Discourse) | Web3 Data Sources (Infura, Quicknode, Alchemy, Coingecko API, Etherscan API, The Graph)
- **Work skills**: Problem solving, Communication (Technical Writing, Documentation, Clear communication), Attention to detail, Project Methodologies (Agile, Scrum, Kanban, Waterfall)
- **Languages**: English (Fluent) | Hungarian (Native)

SOFTWARE ENGINEERING & DATA SCIENCE EXPERIENCE

Data Platform & Operations Engineer | Diligent | Budapest, Hungary | June 2023 – Present

- Enhanced data accessibility by designing and building **serverless data platform infrastructure and pipelines**.

Ensured data flow and retrieval for key stakeholders.

- Overhauled data fetcher logic. Transitioned data API fetchers from VBScript to Python, set up data ingestion to database, and integrated scrapers into cloud infrastructure. Resolved issues with **data quality, performance, and rate limits**. (Python, AWS Lambda, CloudWatch, and MS SQL)
- Built out, improved, and maintained **data operation processes**. Analyzed, improved and systematized **issue resolution pipeline** including resolution time and pipeline reliability and maintainability. (AWS CodePipeline, CloudWatch, EventBridge, SNS, Lambda, and Slack). Orchestrated multi-month backfills.
- Designed **structured logging, notifications, and dashboards** for data and infrastructure monitoring.
- **Reduced code build time by 75%** through refactoring. Expanded CI/CD functionalities to increase build reliability and developer experience. (AWS CodeBuild, Lambda, CodeArtifacts, ECR, bash, and GitHub API)
- **AI ambassador**: Integrated company-drived **AI development workflows** into team SDLC.

Web3 Data Engineer | Aragon DAO | Remote | August 2022 – February 2023

- Developed reporting pipelines for **DAO community and governance analytics**. Retrieved and processed **on-chain and off-chain data** and ensured accurate and timely delivery. (Discourse, Discord, Dework, Dune, Python, web3.py, and pandas)
- Designed and built a **financial oversight dashboard for DAOs**. (Python, pandas, Dash, and Anvil)

Data Scientist & Engineer | Freelancer | Remote | September 2018 – June 2023

- Resolved advanced analytics and data challenges across **Web3, Health, and Energy** sectors.
- Built **analytics pipeline** for crypto arbitrage opportunities. Collected, processed and analyzed on-chain Terra/Cosmos data. (Flipside, Python, pandas)
- Developed a time-series machine learning glucose **forecasting model**. Achieved the prediction accuracy of the market-leading commercial medical devices. (Python, numpy, pandas, scikit-learn)
- Built a **reporting pipeline** from clinical trial data assessing a medical device. Generated actionable insights informing investment decisions. (Python, pandas, matplotlib, seaborn, jupyter)
- Designed a machine learning **feature engineering** evaluation pipeline. (Python, numpy, scikit-learn)

Junior Consultant & Technical Writer | Dorsum | Budapest, Hungary | January 2016 – May 2018

- Supported the development of a **B2B wealth management SaaS** with **regulation analysis, documentation, business proposals, and marketing content**.

EDUCATION

Ph.D. in Science and Technology Studies | The Open University | 2010 – 2016

- Ethnographic research on knowledge and technology in financial innovation | Fieldwork on local currency

Diploma (BA + MA) in Economics | Budapest University of Technology and Economics | 2002 – 2007

- Micro- and macroeconomics, mathematics (calculus, linear algebra), economic statistics, econometrics, optimization | Viability study of digital payment schemes | Specialization in economic analysis | Dissertation on economic growth models

SIDE-PROJECTS: WEB3 DATA ENGINEERING & ANALYTICS

Aave Liquidity Provider TVL Point Tracker | 2024 October – November

- Built a point tracker evaluating **Aave** liquidity providers: https://github.com/nocibambi/aave-lp-point-tracker
- Collected **on-chain data** about assets, liquidity indexes, prices, wallets, and balances. (The Graph, Coingecko, Etherscan, web3.py).
- Processed datasets and calculated points based on **Aave whitepapers** and documentation. (pandas)
- Exposed the points via a **REST API**. (FastApi)

Token Swap Pool/Market Comparison | 2024 August – September

- Built a data collection and ETL pipeline comparing **cryptocurrency token swap** platforms. (pandas, Pydantic)
- Researched and identified reliable **Web3 DEX and CEX data** sources. (CoinGecko, Binance, Dune)
- Developed an ETL pipeline to fetch, parse, transform, and store **cryptocurrency market data**.

Publications: Machines of Trust | Medium | Dev.to