

# Rajeev Paudel

Email: [rajeevpaudel58@gmail.com](mailto:rajeevpaudel58@gmail.com) • Location: Kathmandu, Nepal

## Professional Summary

---

Data Engineering Team Lead with 2+ years building production data pipelines for blockchain analytics, social media intelligence, and media ingestion. Full-stack ownership in practice: ELT design in dbt and Prefect, warehousing in ClickHouse at multi-TB scale, sub-minute latency requirements held in production. Led the data engineering team that shipped 5 products in under a year. I care about data quality, reliability, and pipelines that don't fail silently.

## Professional Experience

---

### Data Engineering Team Lead at 108 Capital

Remote • Apr 2025 – Present

- Led a team of engineers to ship **5 products to production**, owning architecture, deployment, and monitoring for each.
- **EVM On-Chain Indexer (Production)**: Scaled a multi-chain blockchain data platform covering Ethereum Mainnet, Base, BSC, Arbitrum, and Optimism, handling **20 TB+** of on-chain data. Used Kafka stream processing and ClickHouse UDFs for real-time ABI decoding, maintaining **sub-1-minute end-to-end latency**.
- **Socials Intelligence Platform (Production)**: Built a Twitter data pipeline covering account management and ETL orchestration (Airflow + ClickHouse). Deployed **NLP model UDFs** on a distributed GPU-enabled ClickHouse cluster for in-warehouse sentiment analysis and entity extraction; worst-case ingestion delay: 45 minutes.
- **AV Engine**: Built a media transcription pipeline that tracks new YouTube channels and podcast feeds, downloads content, and transcribes with AI. Orchestrated using Plomberry with Celery for async background tasks.
- **Narrative Intelligence Engine**: Built a multi-platform newsletter and article scraper covering Substack and Seeking Alpha, including paywalled content via authenticated browser sessions. Ingested **10 000+ articles** into SurrealDB using Browserless.
- **Reddit Engine**: Built a subreddit and user-tracking ingestion pipeline using SurrealDB as both data lake and warehouse, removing the need for a separate ETL layer.
- Led a **ClickHouse → Trino migration**: designed a benchmarking framework across cluster configurations and sizes and documented the findings for the team.

### Data Engineer at 108 Capital

Remote • Oct 2023 – Mar 2025

- **On-Chain Data Indexer (Build)**: Built a ClickHouse data warehouse from scratch as the sole owner. Designed ETL flows in dbt + Prefect to index EVM blockchain data at **< 1-minute latency**; operating at **multi-TB scale** from day one.
- **Twitter Data Pipeline (Build)**: Built the full pipeline: bot-account provisioning, request orchestration in Airflow, and reverse-engineered Twitter data ingestion into ClickHouse.
- Worked on an ongoing migration to Apache Spark-based ETL.

### Software Engineer — Frontend & Real-Time Comms at Arman Group of Companies

Remote • Jan 2023 – Oct 2023

- Built *Arman Ludo*, an online multiplayer mobile game using Flutter, and owned the real-time communication layer via WebRTC — enabling low-latency peer-to-peer gameplay across players.

## Technical Skills

---

<b>Orchestration</b>	Prefect, Apache Airflow, Celery, Plomberry
<b>Transformation</b>	dbt, SQL, Python
<b>Data Warehouses</b>	ClickHouse (expert), Trino, Snowflake (familiar)
<b>Streaming</b>	Apache Kafka, stream processing, ClickHouse Materialized Views
<b>Databases</b>	SurrealDB (multimodal), PostgreSQL
<b>Languages</b>	Python (primary), Dart/Flutter, Go, SQL
<b>Cloud &amp; Infra</b>	Docker, Linux, distributed cluster management
<b>Real-Time / Other</b>	WebRTC, Browserless (headless scraping), GPU UDFs in ClickHouse
<b>Methodologies</b>	ETL/ELT design, data modelling, pipeline testing, code review

## Education

---

### Bachelor of Engineering — Computer Engineering

Nov 2019 – Apr 2024

*Pulchowk Campus, Tribhuvan University, Nepal*

Ranked **165** / **16,000+** applicants

**Relevant Coursework:** Data Structures & Algorithms • Database Systems • Distributed Systems • Data Warehousing & Business Intelligence • Big Data • Software Engineering

## Key Highlights

---

- **Scale:** Designed and operated pipelines handling **20 TB+** of blockchain data across 5 EVM chains in production.
- **Latency:** Built and held **sub-1-minute** on-chain to warehouse latency; 45-minute SLA for social media ingestion.
- **Leadership:** Grew into a team lead role and delivered 5 production products within the first year of leading the engineering team.
- **Data Quality:** Built in-warehouse NLP pipelines (sentiment, entity extraction) running on distributed GPU ClickHouse clusters.