Production-Grade Data Unification Pipeline for Fello

Transforming Raw Research into Actionable Business Intelligenc

Fello

Adarsh Badjate
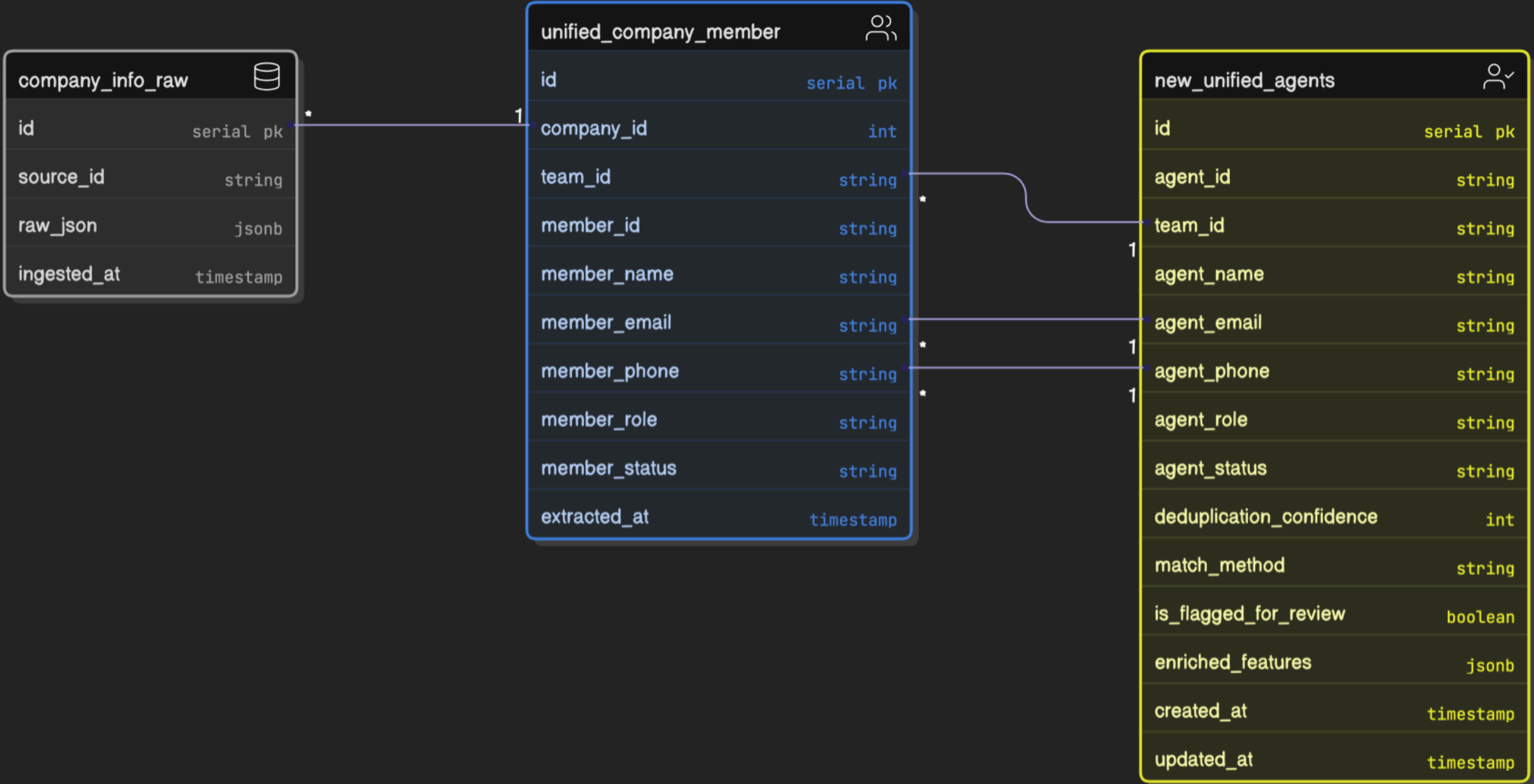15 Feb 2026

The Problem: From Raw Data to Master Data

The core challenge was to unify two distinct data sources into a single, reliable master dataset.

Source 1: Raw Research Data (`bronze.company_info_raw`)

Source 2: Master Agent Data (`public.new_agents`)

Goal : Create a scalable, automated, and production-ready pipeline to produce a clean, unified `public.new_unified_agents` table.

Fello

Schema Design

Key Decision 1: Flattening JSON at the Core

The most critical architectural decision was how to handle the nested JSON data.

Tradeoff: Direct JSON queries vs. a structured Silver Layer.

Decision: We chose to **flatten the JSON** into a relational `unified_company_member` table.

Why?
1. **Performance**
2. **Simplicity**
3. **Maintainability**

Fello

Matching hierarchy (vertical ladder design)

Matching Keys

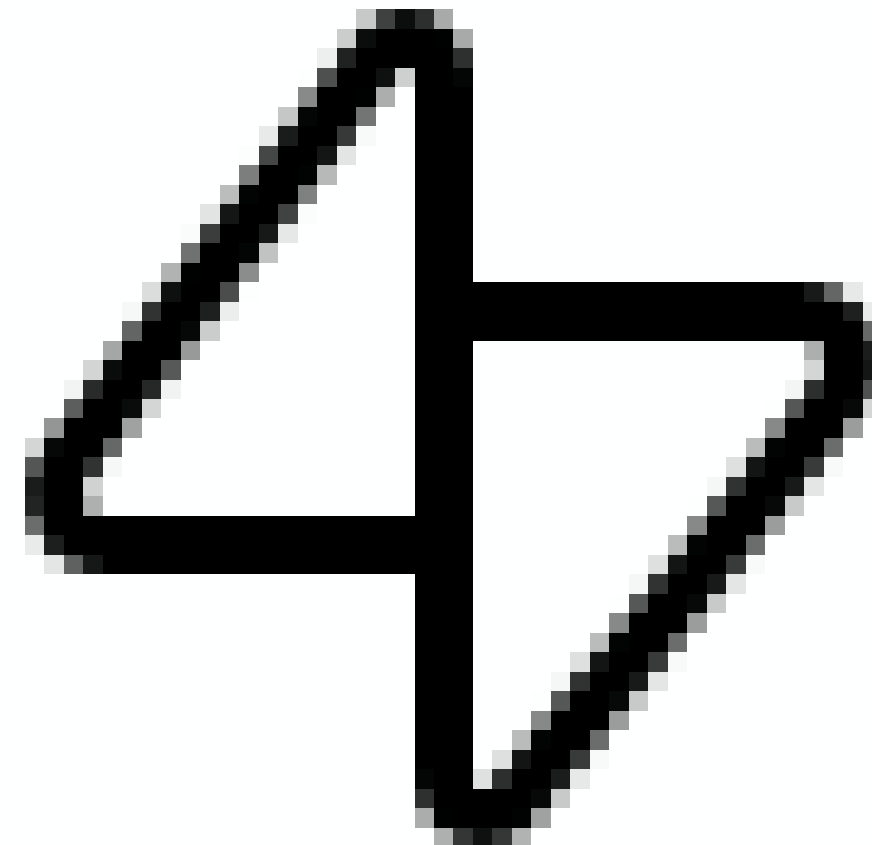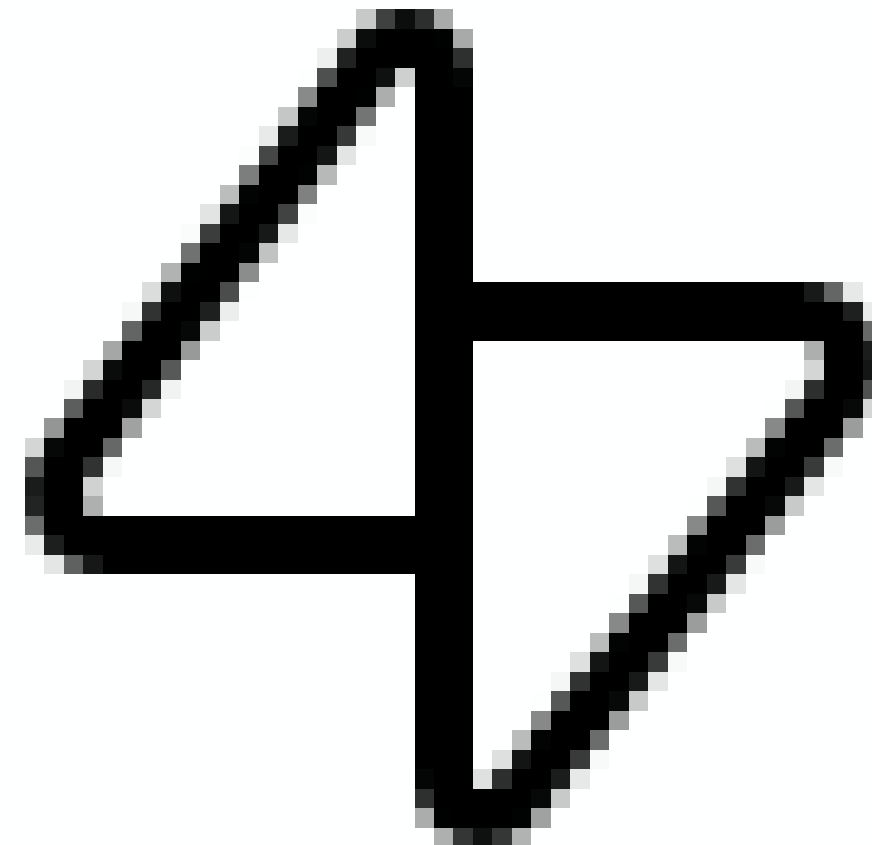|		
Email + Team	Confidence: 100%	✓ Fully automated ✓ Deterministic ✓ Highest reliability
Phone + Team	Confidence: 90%	✓ Strong identifier ✓ Automated match
Phone + Team	Confidence: 90%	✓ Strong identifier ✓ Automated match
Exact Name + Team	**Confidence: 80%**	✓ Structured comparison ✓ Automated but logged
Fuzzy Name + Team	Confidence: 60%	⚠ Requires manual review

Fello

Technologies used

01 Supabase

02 SQL

03 Python

04 N8N

Technologies used

01 Supabase

02 SQL

03 Python

04 N8N

Fello

Deployment Pipeline: Automation & Reliability

python deploy.py --full

✓ **Idempotent – Safe to re-run**
✓ **Transactional – Auto rollback on failure**
✓ **Comprehensive Logging – JSON audit reports**

db_manager

sql deployer

data_loader

Deployment Pipeline: Automation & Reliability

python deploy.py --full

✓ **Idempotent — Safe to re-run**
✓ **Transactional — Auto rollback on failure**
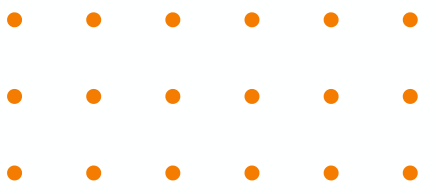✓ **Comprehensive Logging — JSON audit reports**

db_manager

sql deployer

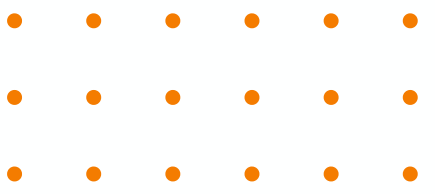data_loader

Data Governence

- **Row Level Security**
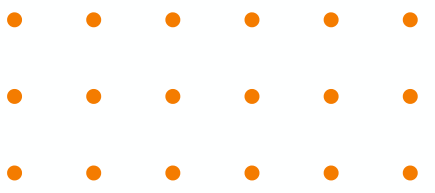- **Supabase Edge Functions**
- **DATA API**

Execution

What I'd Do Differently with More Time

Enhance the platform with an AI-governed data access layer, full CI/CD automation, advanced intelligent matching, and a structured, queryable data dictionary

FUTURE

GOALS

Fello

Thank You