GraphGeeks

kùzu

Getting Started with BAML and Kuzu: Create graphs from unstructured data using LLMs

github.com/graphgeeks-lab/baml-for-graph-101

PRESENTED BY

Prashanth Rao

AI Engineer, Kùzu Inc.

GraphGeeks Workshop

🌎 Online | 20 May 2025

Problem statement

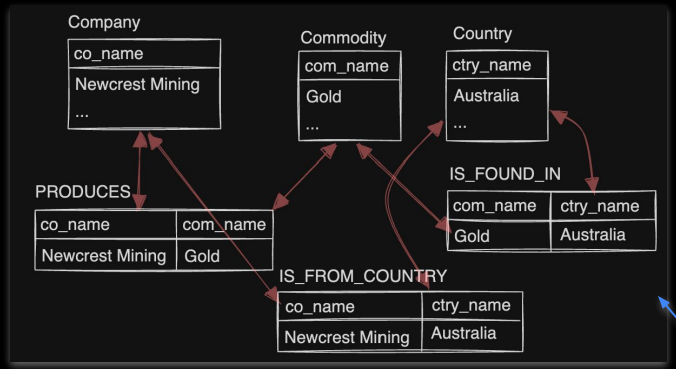
Imagine you are a developer at a financial asset company, with access to a news API feed that updates daily with articles about the various mergers and acquisitions occurring in the mining industry.

Tasks

- Answer questions about connected data (tree structures with hierarchical information)
  - Which companies merged with or acquired other companies?
- Increase the level of autonomy in tracking/analyzing M&A events
  - Signal alert bot
  - Graph RAG-based Q&A system

Data models are interchangeable

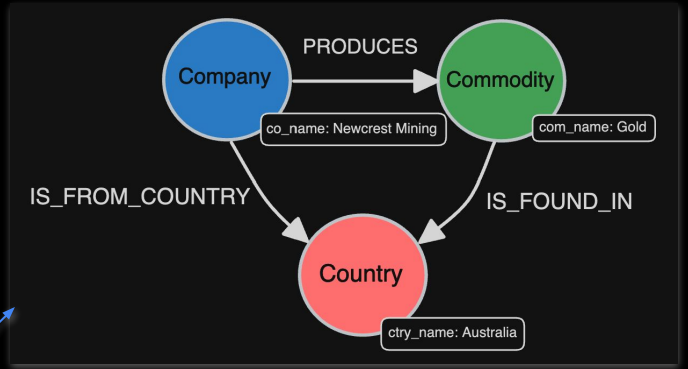
Relational model



Document model

```
[
  {
    "company": "Newcrest Mining",
    "commodities": ["Gold"],
    "country": "Australia"
  },
  {
    ...
  }
]
```

Property Graph model



Tools used in this workshop (code available)

LLM pipeline

An expressive language to get structured outputs from LLMs

boundaryml.com

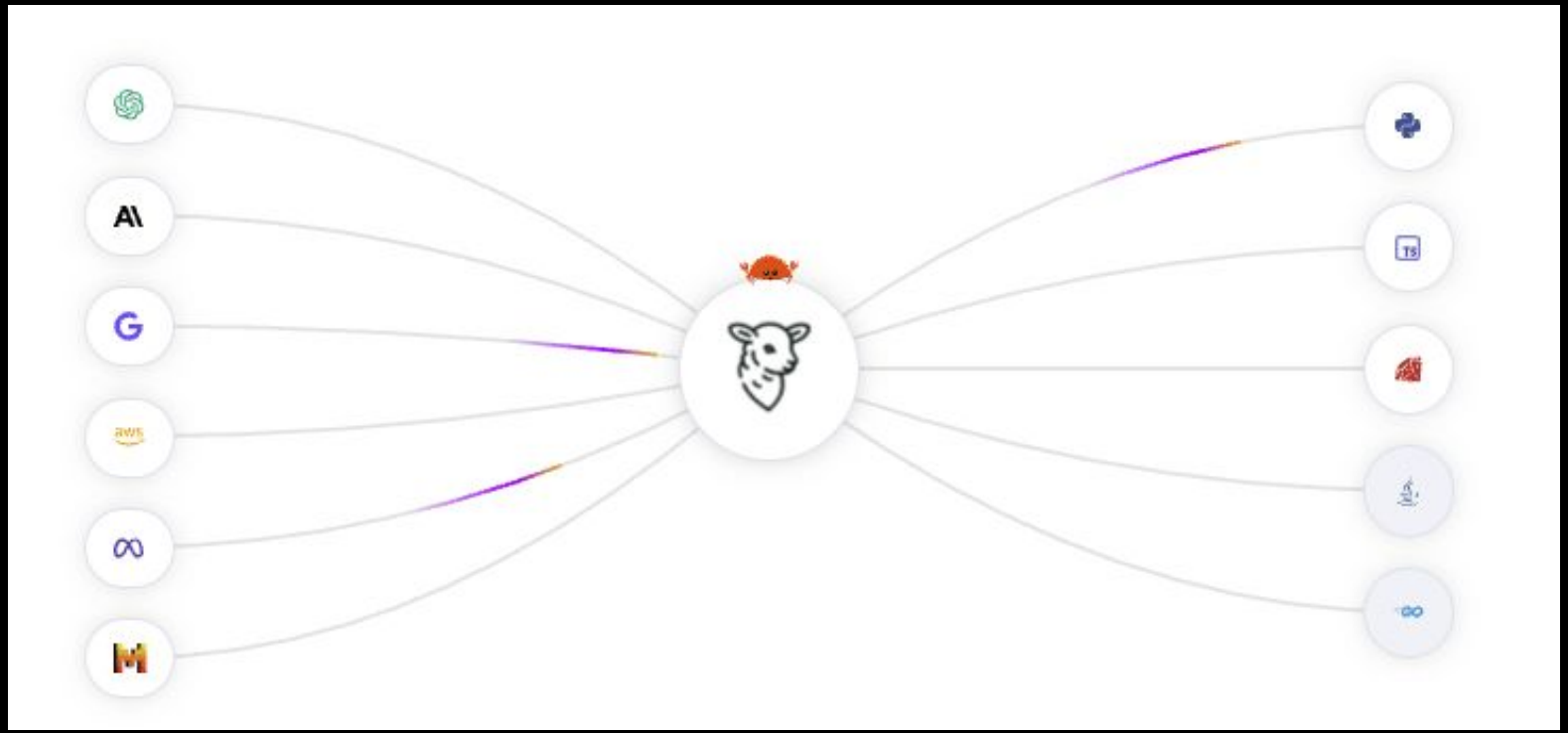
Graph pipeline

Embedded property graph database with a Cypher query interface

kuzudb.com

See the code and data here: https://github.com/graphgeeks-lab/baml-for-graph-101

BAML: Connect any LLM to apps in any language

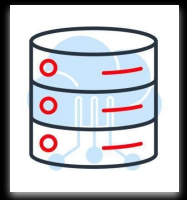


GitHub: github.com/boundaryml/baml

Kuzu: Simplicity without compromising performance

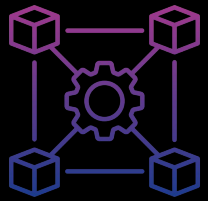
🔥 + 🚀

Performance & scalability



In-process (easy to set up & deploy)

kùzu



Usability & interoperability

MIT license

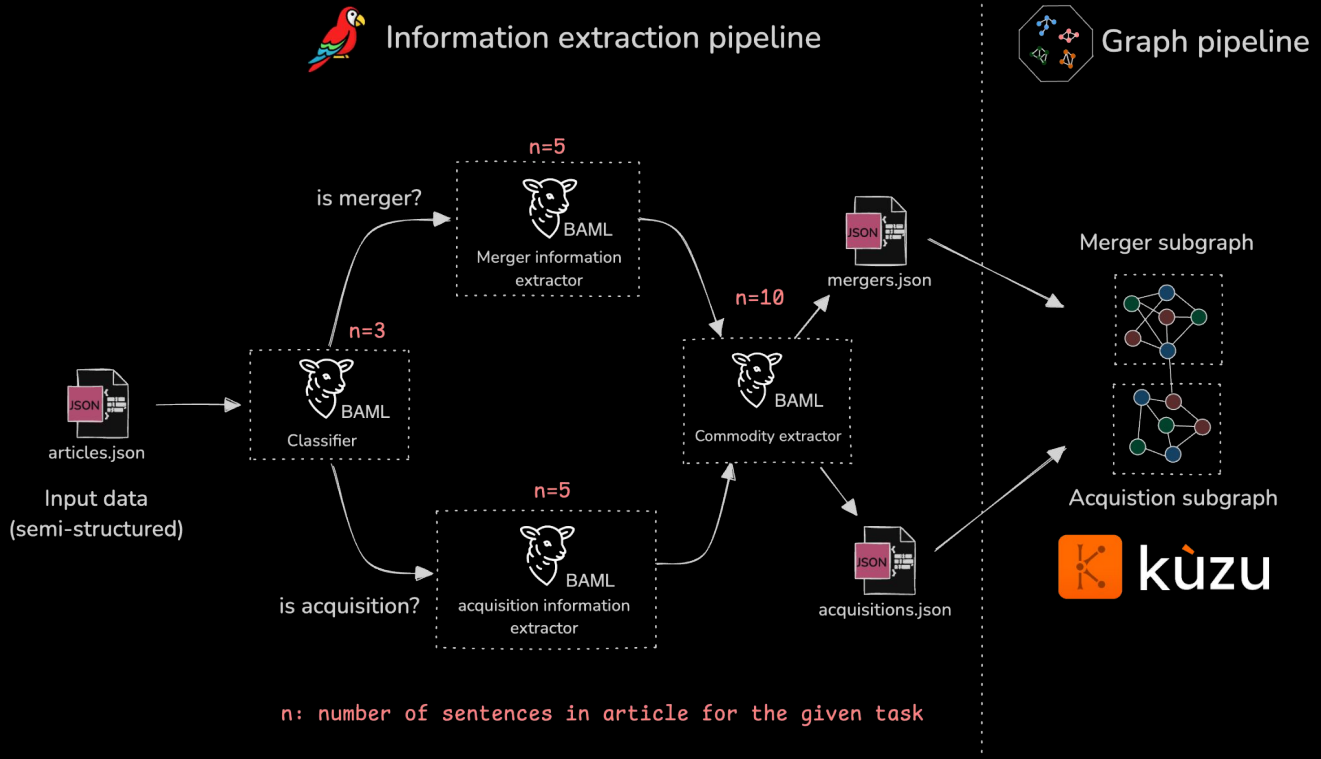
Open source

GitHub: github.com/kuzudb/kuzu

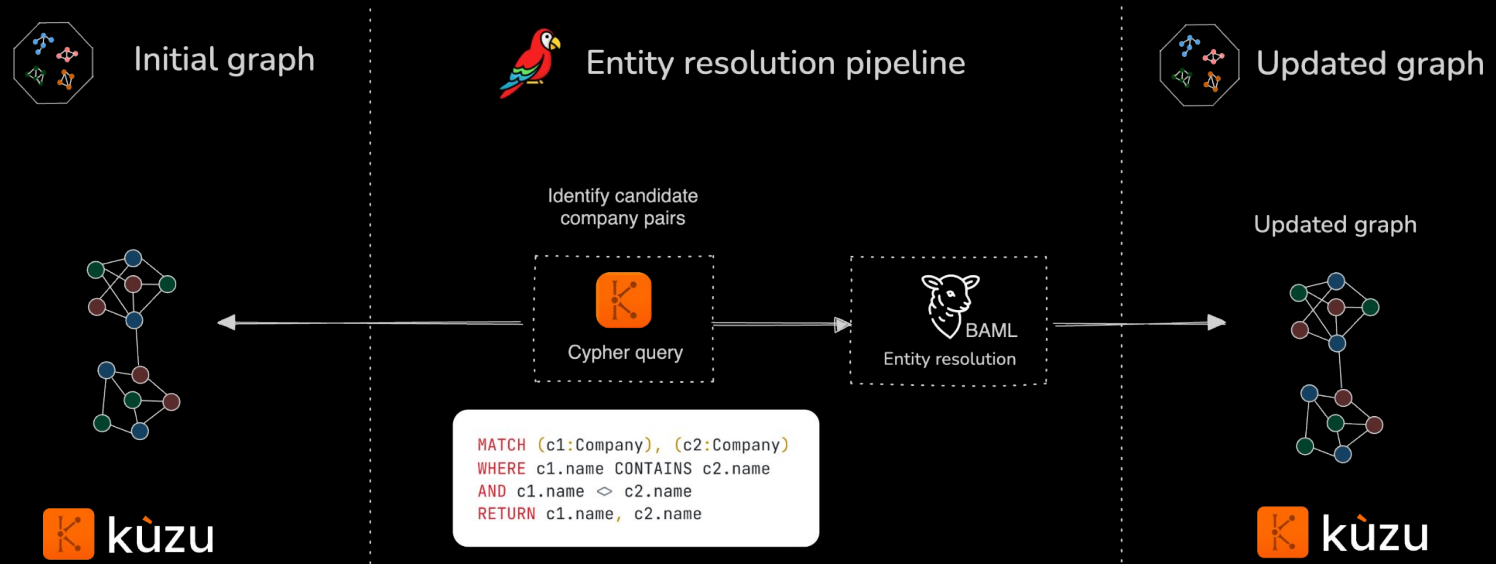
Sample data


```
[
    {
        "id": 1,
        "url": "https://www.australianmining.com.au/elevra-lithium-a-defining-moment-in-north-american-lithium/",
        "title": "Elevra: ‘A defining moment in North American lithium’",
        "date": "2025-04-09",
        "content": "The company to be created through the merger of ASX-listed Sayona Mining and Piedmont Lithium... "
    },
    {
        "id": 2,
        "url": "https://www.mining-technology.com/news/newcrest-mining-pretium-resources-acquisition/",
        "title": "Newcrest Mining completes acquisition of Pretium Resources",
        "date": "2022-03-09",
        "content": "Australia’s Newcrest Mining has closed the acquisition of Pretium Resources, which owns ... "
    },
  ...
]
```

Stage 1: Create initial graph



Stage 2: Improve graph quality



Learning material

Intro to Cypher

Cypher is SQL-like

SQL	Cypher
`FROM`	`MATCH (a:Company)-[t:ACQUIRED_BY]->(b:Company)`
`WHERE`	`WHERE a.name = “Newcrest Mining”`
`SELECT`	`RETURN b.name`

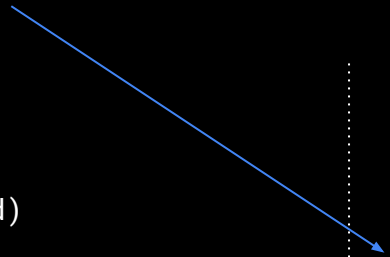
Specialized syntax & functions for graphs


```
MATCH p = (a:Company)-[t:*1..4]-(b:Commodity)
WHERE a.name = “Newcrest Mining”
AND b.name = “Gold”
RETURN p
```

Grouping in Cypher is implicit

There is no `GROUP BY` clause in Cypher

- Implicitly done based on the contents of the `RETURN` clause



SQL

```
SELECT co.name, COUNT(c.id)
FROM Company c
JOIN PRODUCES r ON c.id = r.company_id
JOIN Commodity co ON r.commodity_id = co.id
GROUP BY co.name;
```

Cypher

```
MATCH (c:Company)-[r:PRODUCES]->(co:Commodity)
RETURN co.name, count(c)
```

Takeaways

- BAML provides the following key benefits:
  - Powerful & expressive language to express prompting logic
  - Type system for LLMs
  - Fixes LLM outputs to ensure quality structured generation
  - Unit tests for prompts!
- Kuzu provides the following key benefits:
  - In-process, on-disk persistence layer for property graphs
  - Cypher query language for concise joins (for humans and LLMs)
  - Easy data ingestion from various sources
  - Fast & scalable for large graphs

Chat with other graph and LLM developers on Discord!

BAML Discord

Kuzu Discord