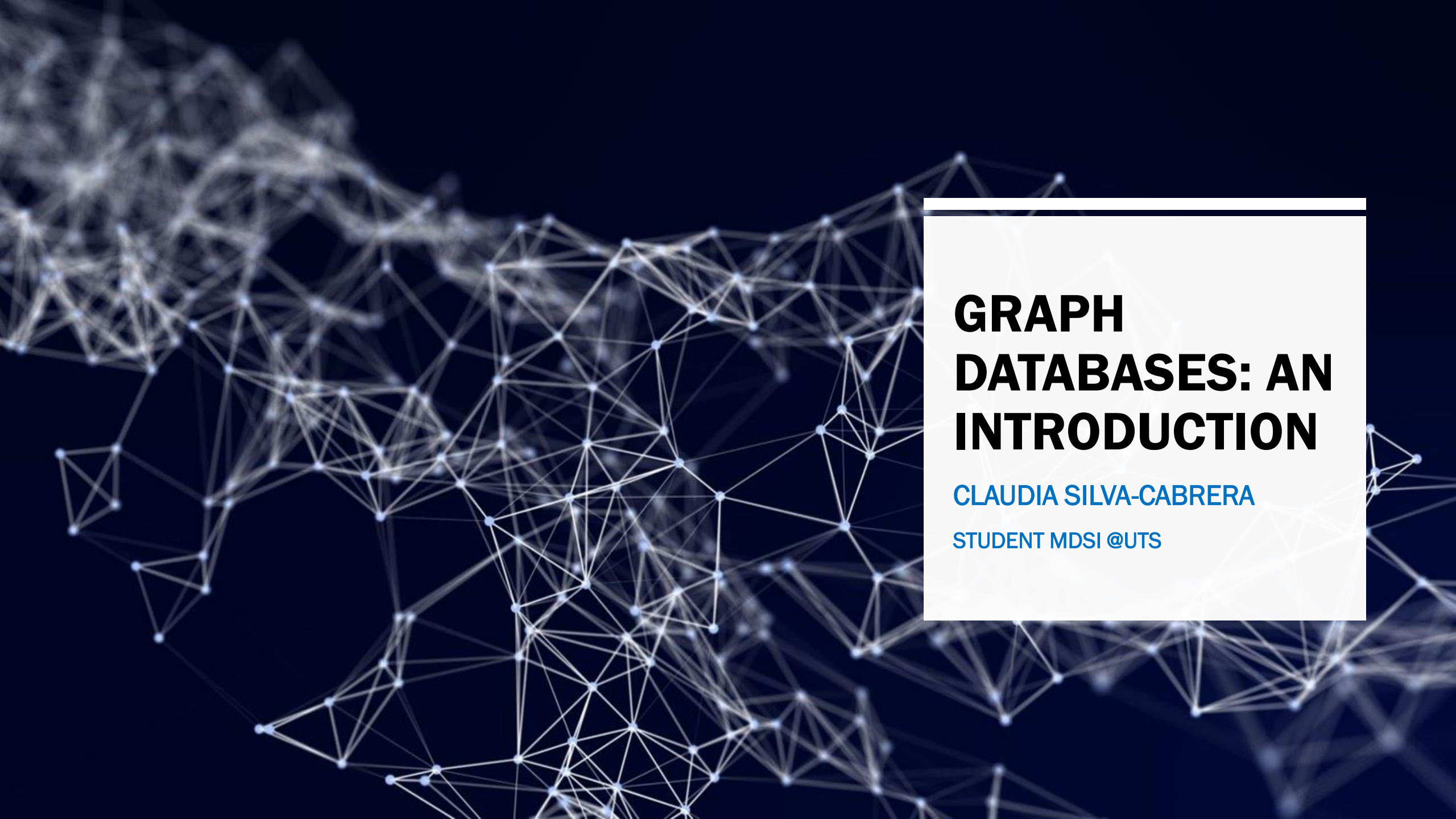


GRAPH DATABASES: AN INTRODUCTION

CLAUDIA SILVA-CABRERA

STUDENT MDSI @UTS

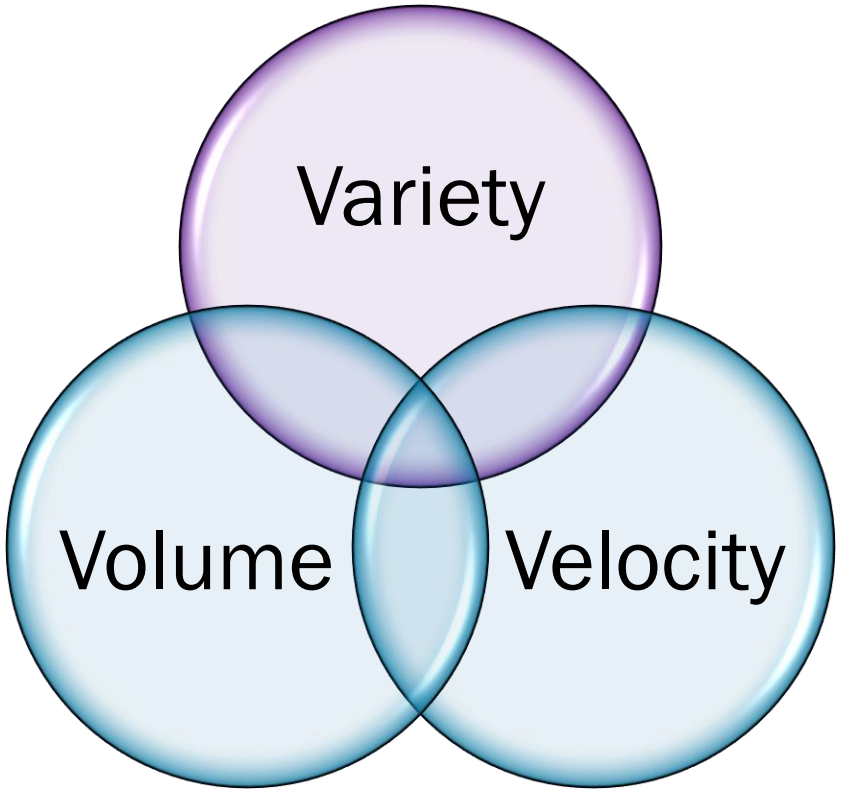
CONTENTS

01

Why now?

Boom of Big Data

- Structured
- Unstructured
- Semi-structured



- Terabytes
- Petabytes
- Exabytes

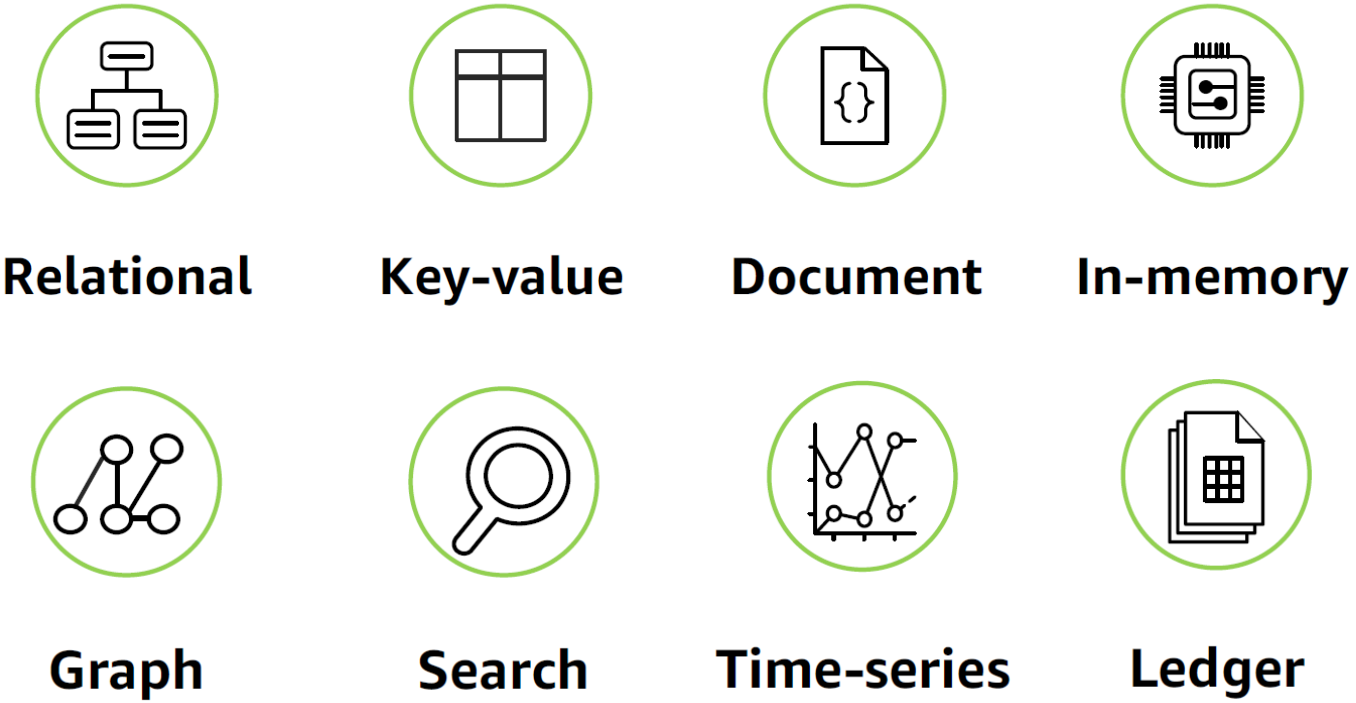
- Batch
- Real-time
- Streaming

02

One size does not fit all

“Purpose-built era”

Find the right database type for the right job



Credits: Amazon AWS Databases “Enter the Purpose-Built Database Era”

03

Who is using Graph Databases

In 2015 Facebook had more than 1 million of users, 100 billion of ratings and millions of items. They extended **Apache Giraph** as an underlying framework to distribute iterative and graph processing.

Real-time recommendation engine. Keeping the **context of the information** to looking for products.

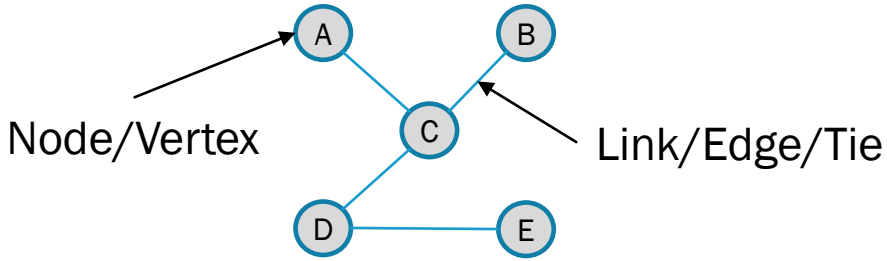
Graph that captures biological knowledge linked to research. There are half a billion relationships in the knowledge graph, and it is expected to triple that number as data is added.

04

Graphs: Elements and Properties

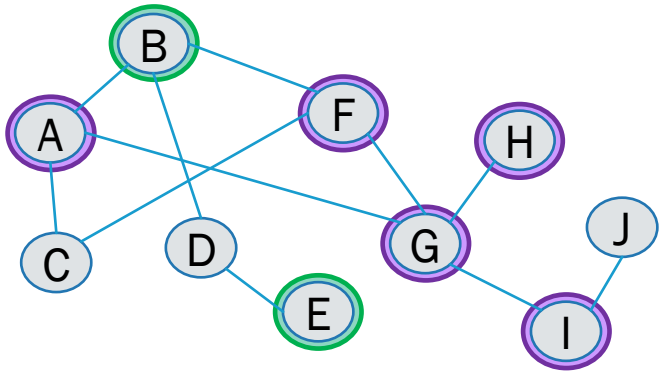
A network is equivalent to a graph

Elements



Main Properties

- **Adjacency:** If a vertex is next to another.
- **Connectivity:** If there exists a path connecting two specific vertices.

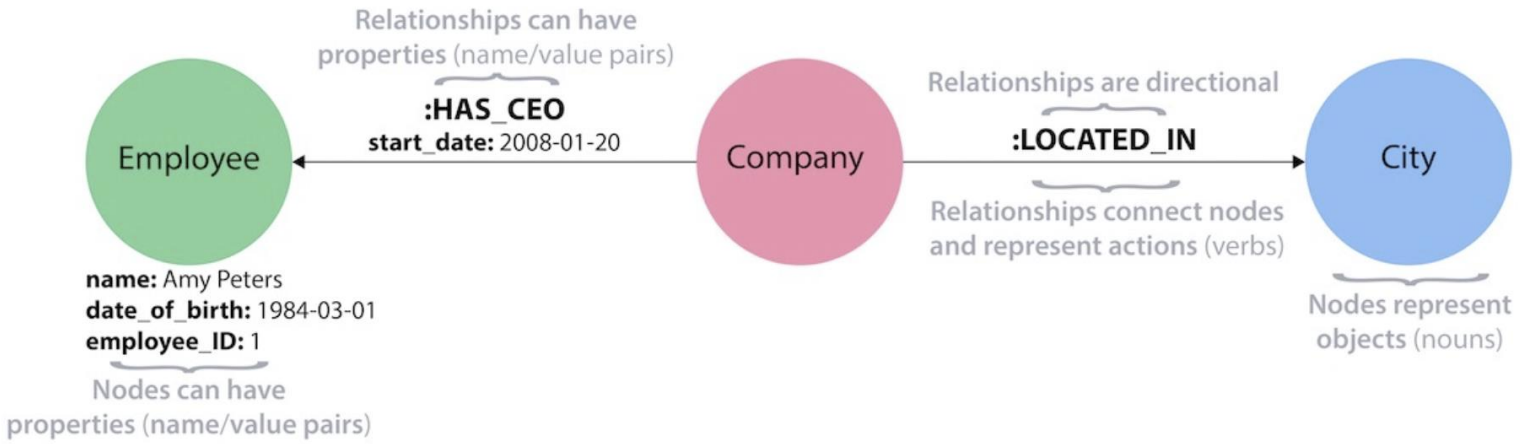


Credits: UTS subject “Social Information and Network Analysis”

05

What is a Graph Database?

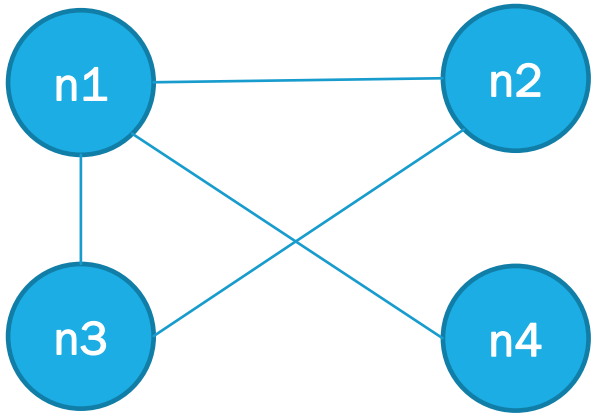
- It is a structure that leverages the concept of **relationship** among entities.
- Uses a mathematical representation of a **graph** with three main elements:
  - NODES
  - EDGES: the relationship
  - PROPERTIES



Credits: NEO4J “What is a Graph Database?”

06 Elements of a Graph Database

- **Graph Storage:** Usually is the native graph structure:



Matrix

	n1	n2	n3	n4
n1	0	1	1	1
n2	1	0	1	0
n3	1	1	0	0
n4	1	0	0	0

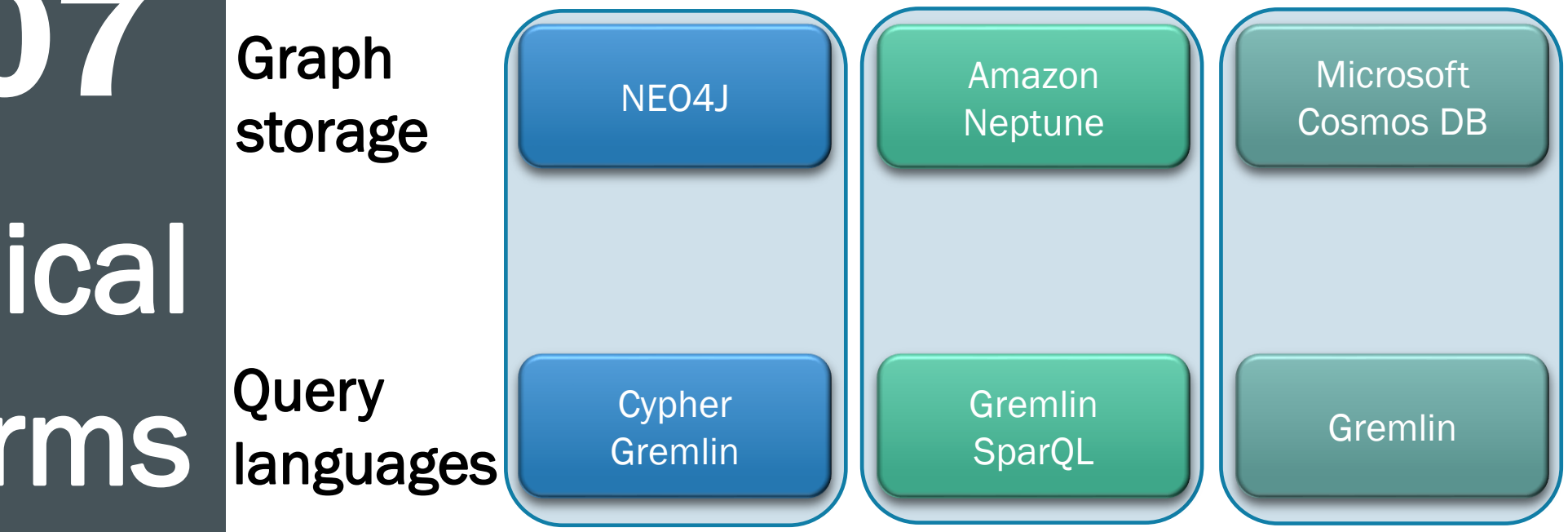
List

n1 → (n2, n3, n4)
n2 → (n1, n3)
n3 → (n1, n2)
n4 → (n1)

- **Graph Processing Engine:** Algorithms and queries supported by the graph storage.

Credits: UTS subject “Social Information and Network Analysis”

07 Technical platforms



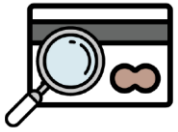
08 Business Use Cases

Everywhere where does exist a *relationship* among entities.

Social networking

Recommendations

Knowledge graphs



Fraud detection

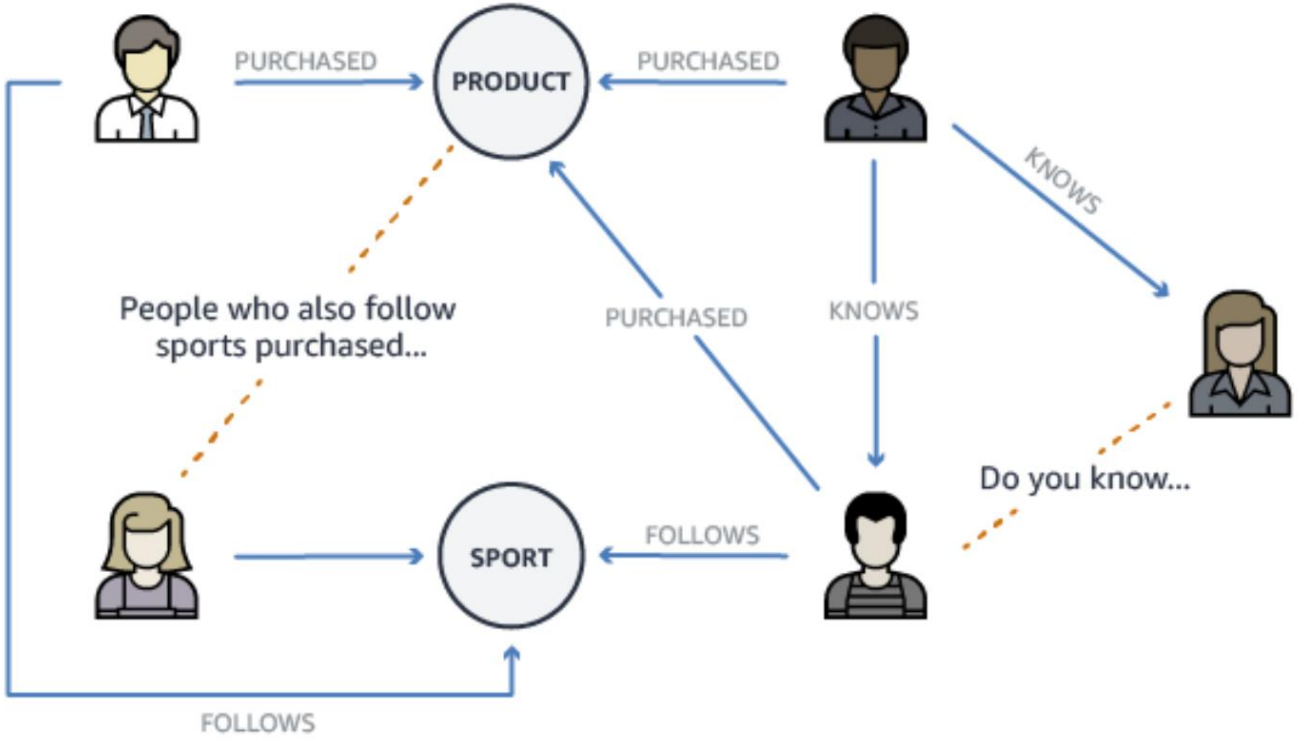
Life Sciences

Network & IT operations

Credits: AWS “Databases on AWS: How To Choose The Right Database”

09

Example: Recommendation System



Credits: Amazon Neptune “Overview: Recommendation Engines”

09.1

Recommendation System

Relational Model

User

UserID	User	Address	Phone	Email	Alternate
1	Alice	123 Foo St.	12345678	alice@example.org	alice@neo4j.org
2	Bob	456 Bar Ave.		bob@example.org	
...	...	...	...	...	...
99	Zach	99 South St.		zach@example.org	

Order

OrderID	UserID
1234	1
5678	1
...	...
5588	99

LineItem

OrderID	ProductID	Quantity
1234	765	2
1234	987	1
...	...	...
5588	765	1

Product

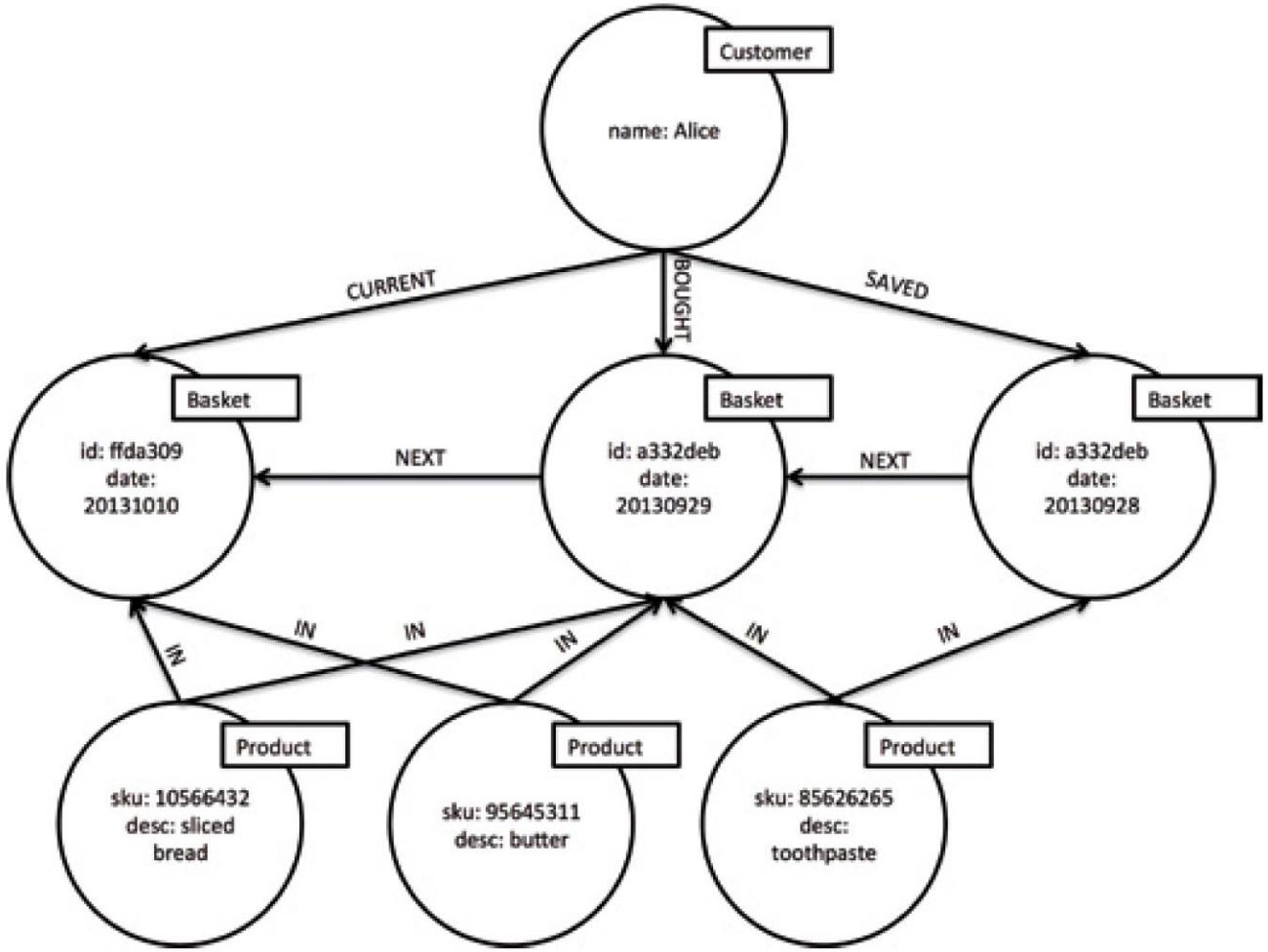
ProductID	Description	Handling
321	strawberry ice cream	freezer
765	potatoes	
...	...	
987	dried spaghetti	

Credits: NEO4J “Graph databases for beginners”

09.2

Recommendation System

Graph Model



Credits: NEO4J “Powering Real-Time Recommendations with Graph Database Technology”

09.3

Recommendation System

Graph Query

Find 5 popular products across the store:

```
MATCH (customer:Customer)-[:BOUGHT]->(:Basket)
      <-[:IN]-(product:Product)
RETURN product, count(product)
ORDER BY count(product) DESC LIMIT 5
```

Find popular products in her social network:

```
MATCH (customer:Customer {name: 'Alice'})-[:FRIEND*1..2]->(friend:Customer)
WHERE customer <> friend
WITH DISTINCT friend
MATCH (friend)-[:BOUGHT]->(:Basket)<-[:IN]-(product:Product)
RETURN product, count(product)
ORDER BY count(product) DESC LIMIT 5
```

Credits: NEO4J "Powering Real-Time Recommendations with Graph Database Technology"

Comparison execution time between a RDBMS and NEO4J

- **Query:** find friends-of-friend connections to a **depth of five degrees**. Dataset with 1M people each with 50 friends on average.

Depth	RDBMS execution time(s)	Neo4j execution time(s)	Records returned
2	0.016	0.01	~2,500
3	30.267	0.168	~110,000
4	1543.505	1.359	~600,000
5	Unfinished	2.132	~800,000

Credits: NEO4J “Graph databases for beginners”

10 Where graphs fit

Handle this situations:

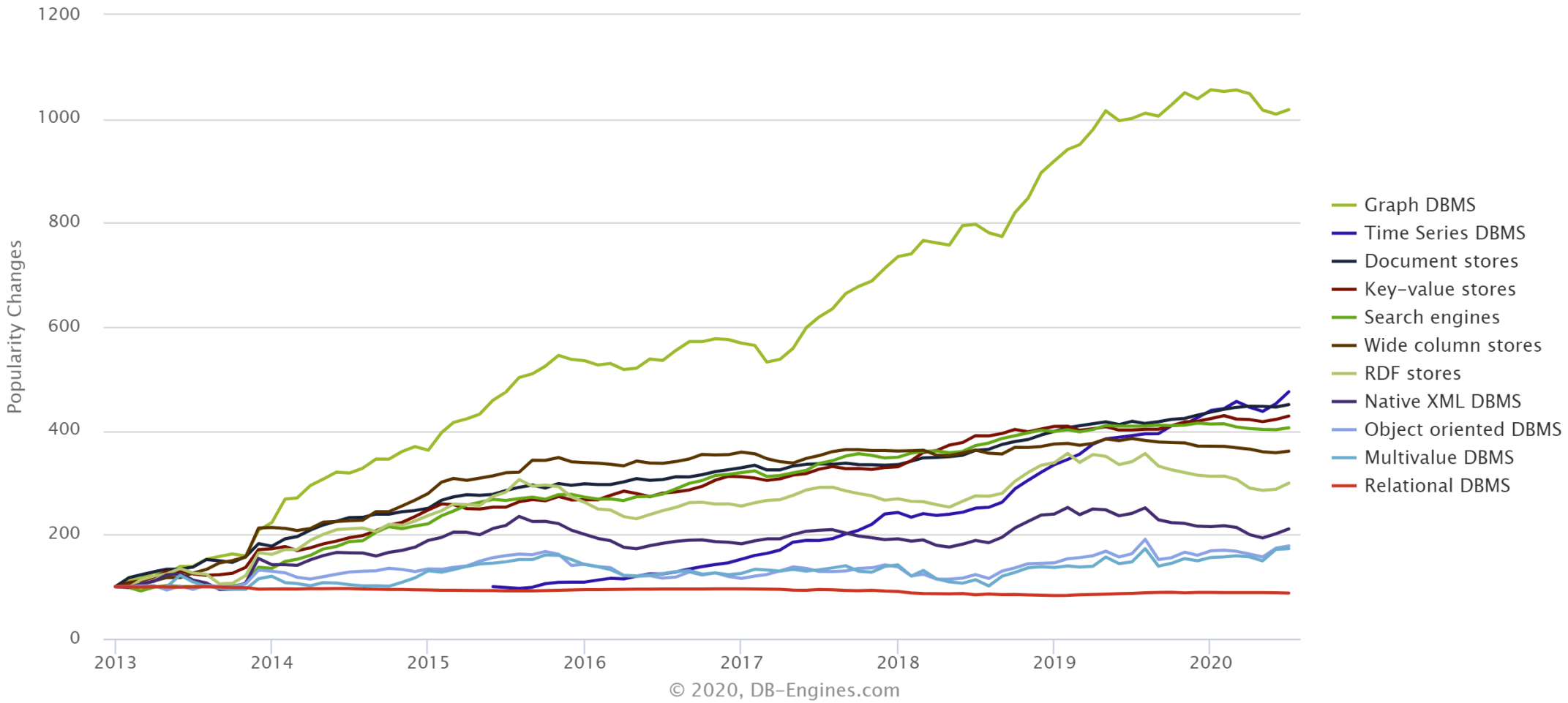
- Frequent or online schema changes
- Establish quick relationships between many different types of data
- Increasing number of connections

Not designed for:

- Applications that do not traverse or query relationships
- Handling queries that span the entire database

Credits: Amazon AWS Databases “Enter the Purpose-Built Database Era”

DBMS Popularity broken by database model

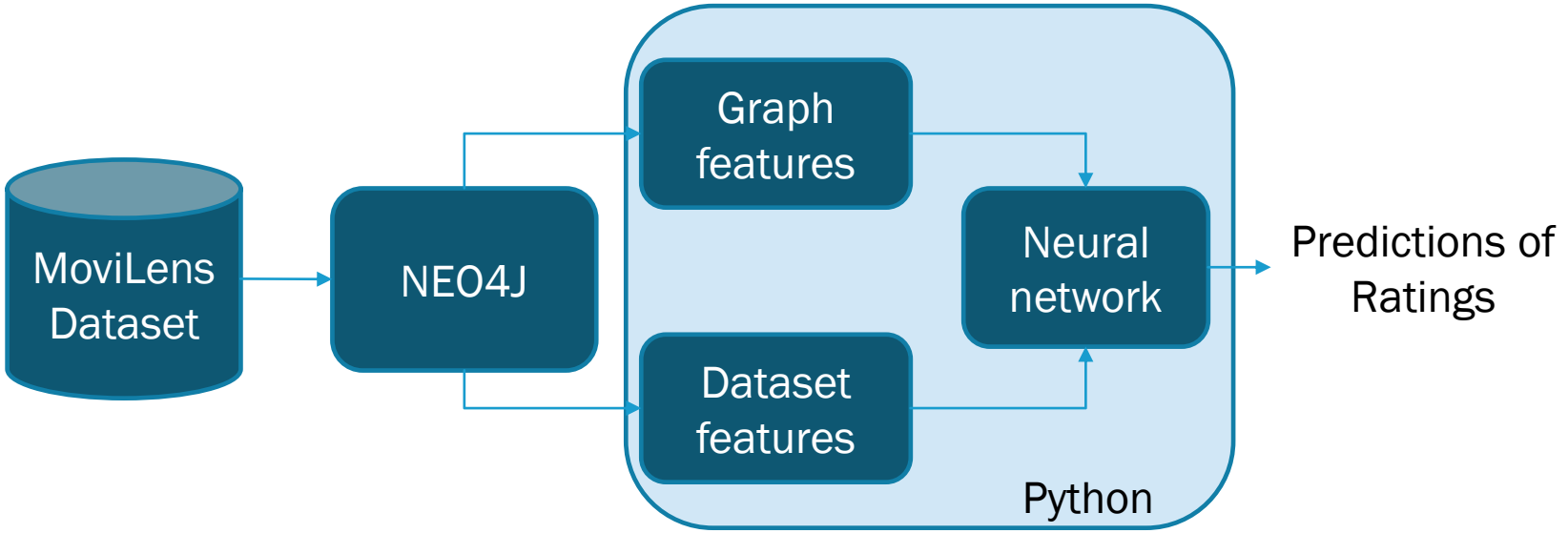


Credits: DB-Engines

11

Demo

Using graph features for making predictions



https://github.com/crsilvac/GraphTheory

12

Credits

NEO4J:

https://neo4j.com/

AMAZON DATABASES:

https://aws.amazon.com/products/databases/

AZURE COSMOS DB:

https://docs.microsoft.com/en-us/azure/cosmos-db/graph-introduction

DB-Engines:

https://db-engines.com/en/ranking_categories

FACEBOOK *Recommending items to more than a billion people:*
https://engineering.fb.com/core-data/recommending-items-to-more-than-a-billion-people/

THANK YOU

QUESTIONS?