

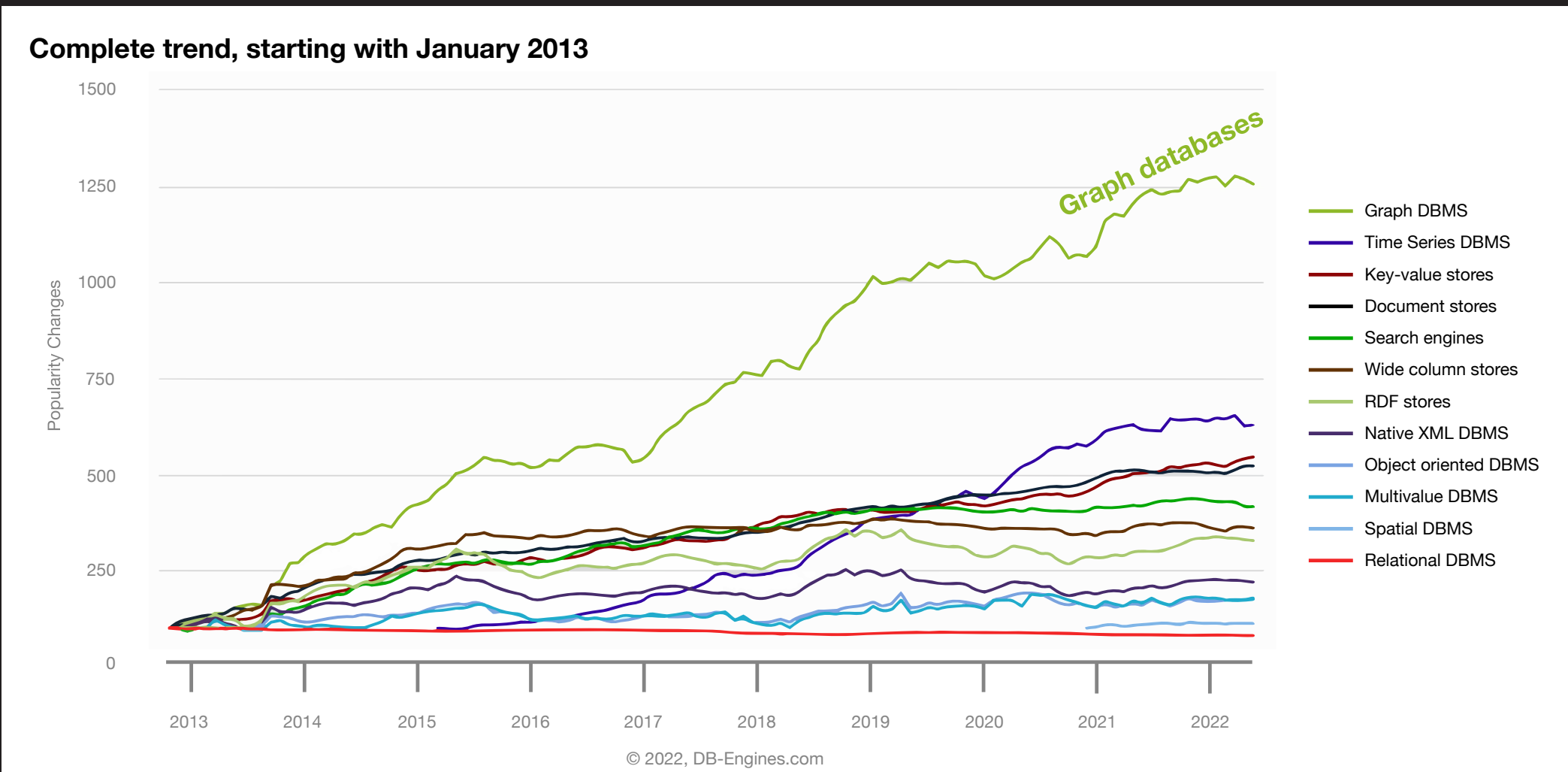
The Table has Turned

■ Catharina Dahlbo & Anna Berg, Capish, Sweden ■
catharina.dahlbo@capish.com • anna.berg@capish.com



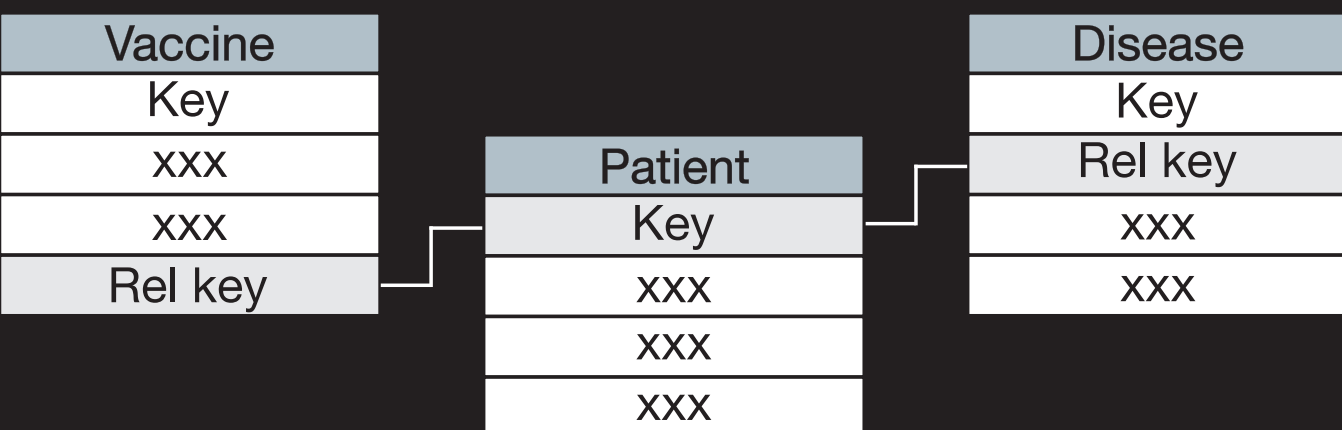
Database technologies have been expanding over the last decades and there are now several options to choose between. Still the relational database constitutes the most common technology, but recently the NoSQL databases, including the graph databases, have gained a lot in popularity.

Choosing between a graph database and a relational database boil down to how the database will be used, since both types offer advantages.



Graph databases continue to grow in popularity. Source:DB-Engines.com

Relational DB - For Predefined Analysis



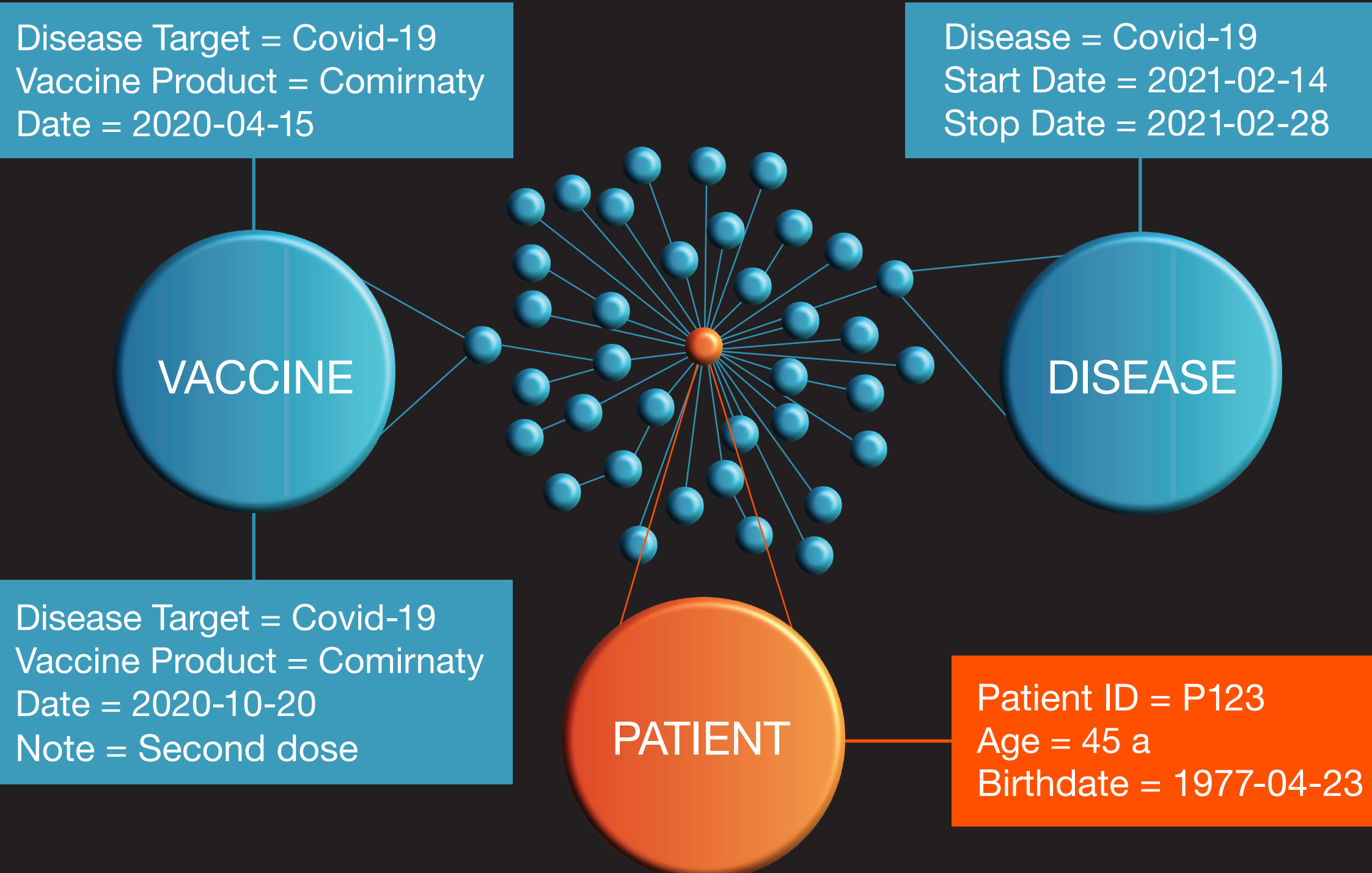
THE RELATIONAL DATABASE has been and still is the most common way to store and analyze data in the pharma industry. Two examples are Oracle and MySQL.

A relational database is based on tables with rows and columns and requires a predefined and carefully modeled schema. To model a relationship, another table must be added to the schema. If there are new relations or variables, the schema would have to be changed accordingly.

The need for a strict data model leads to that developers first must spend a lot of time to get the model right and then spend a long time adjusting it if the business use-case changes. The relational database is therefore designed for updates of data values rather than updates of the data structure.

The well-structured and strict schema makes the relational databases ideal for situations where the same variables are used for a long time, e.g., accounting and bank transactions.

Relational databases are also often designed for high data integrity, making them suitable for financial or defense data.



Which patients have had Covid-19 and have taken two doses of Covid-19 vaccine?

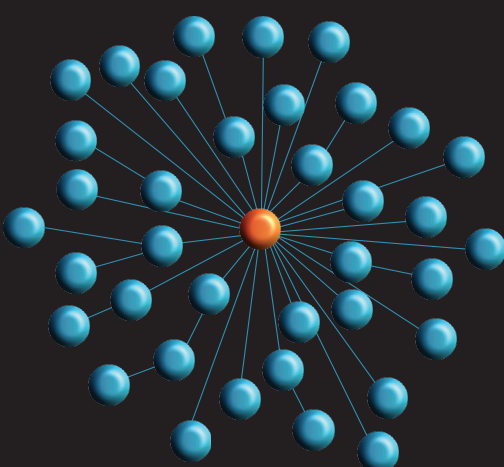
Relational DB – Create tables

- 1. Join and limit tables to Covid-19 disease & vaccines.
- 2. Group table to count vaccine doses/patient.
- 3. Limit the table to two doses.
- 4. Create new tables for further exploration of these patients' data.

Graph DB – Create subgraph

- 1. Limit the patient nodes to those with at least one Covid-19 disease node and two Covid-19 vaccine nodes.
- 2. Continue refining the subgraph for further exploration of these patients' data.

Graph DB - For Data Exploration



IN THE PHARMA industry, graph databases have been gaining in interest for several years, but their benefits are still not widely known. Examples include Stardog, Neo4j and Capish Reflect.

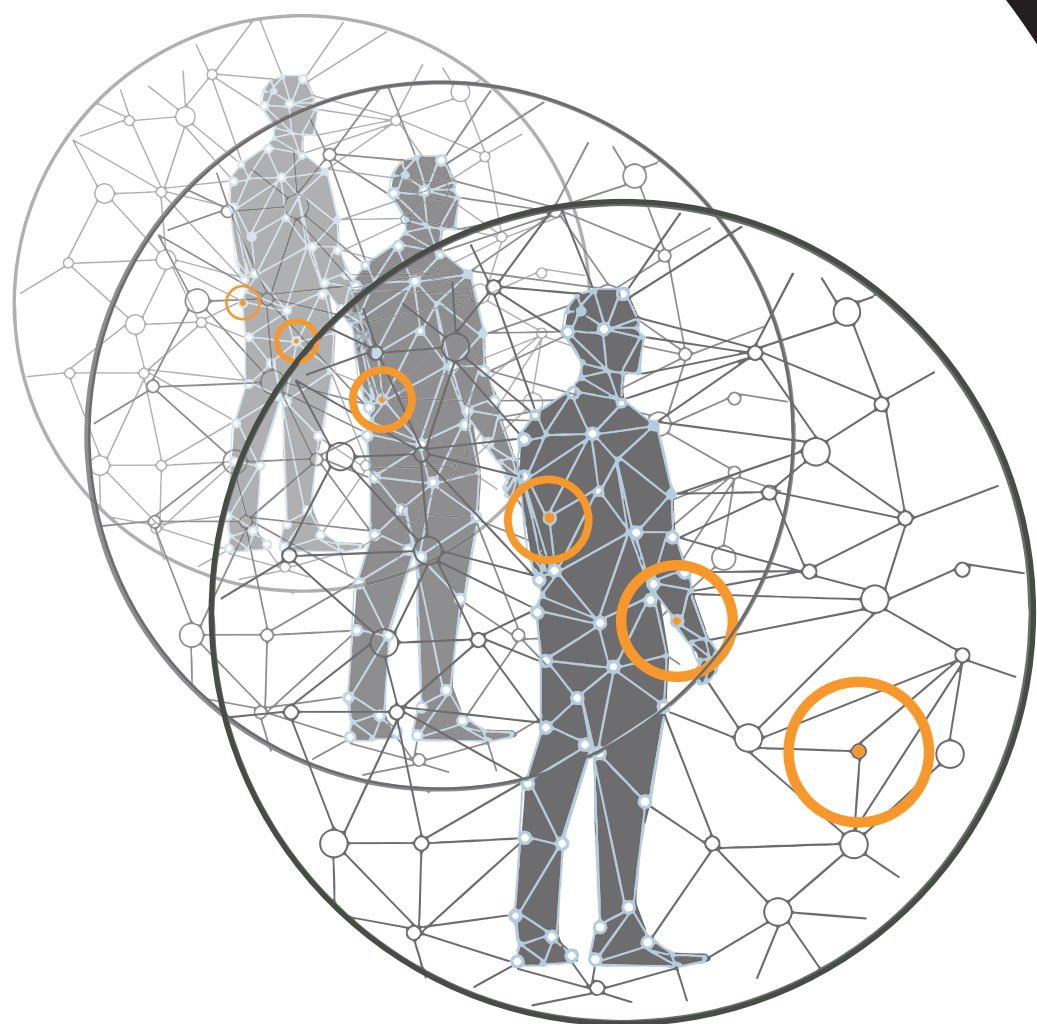
Graph databases store data as nodes (people, places, drugs, etc.) and relationships that connect the nodes. The graph database does not require a predefined data structure, instead nodes and relationships can be added when needed.

The flexible modelling is optimal for cases when there is a need for repurposing of data, to answer new questions and broaden the original analysis scope, e.g., when integrating and harmonizing data from different sources.

The explicit relationships make it possible to traverse the graph to find patterns and similarities within the data, both exploratory and by using machine learning algorithms. This is useful for a wide range of data science use cases, from customer recommendations to finding similar patients.

Two common graph database techniques are the RDF (Resource Description Framework) and the property graph databases. A property graph stores more information within each node and/or relationship, which makes the graph faster to traverse and easier to interpret.

Relational Databases		Graph Databases	
Tables with columns and rows	Format & Model	Graphs with nodes and edges	Use Cases
Implicit relations using table keys		Explicit relations using edges	
Rigid data model (each row has the same columns)		Flexible data model (each node can contain different properties)	
Complex joins between tables required		No joins are required	
Predefined and flat data (e.g., financial data)	Use Cases	Varying and interconnected data (e.g., medical data)	Data science (exploration)
Data entry (with transactions)		Integration and harmonization	
Predefined analysis			



DATA FROM MULTIPLE sources can be curated, integrated, and stored in a property graph database to improve accessibility and usability. Capish has refined the property

graph database by using an ontology-based information model and by using innovative indices to make the querying process faster and more intuitive.