TigerGraph

What's a Graph Database and Why Should You Care?

We Use Graphs Every Day

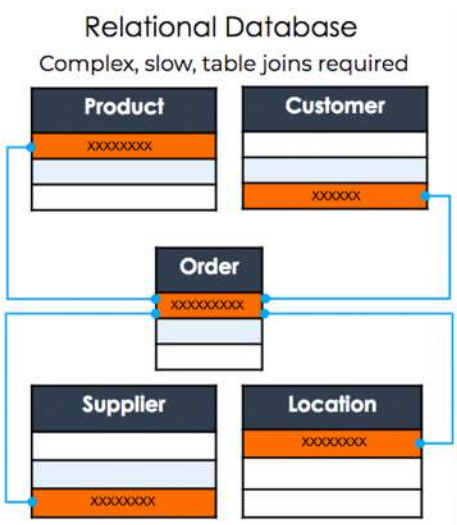
You use graph databases every day and probably don't know it. Every time you visit LinkedIn and see first-, second-, or third-degree connections, you're getting results from the social networking site's professional network graph built on a graph database. Facebook, Instagram, and Twitter all use graph databases and analytics to understand how users relate to each other and connect them with the right content. Every time you do a Google search, you're tapping into Google's Knowledge Graph. Those product recommendations on Amazon -- "people who bought this item also bought…" or "these items are often bought together"? That comes from a graph analytics query too.

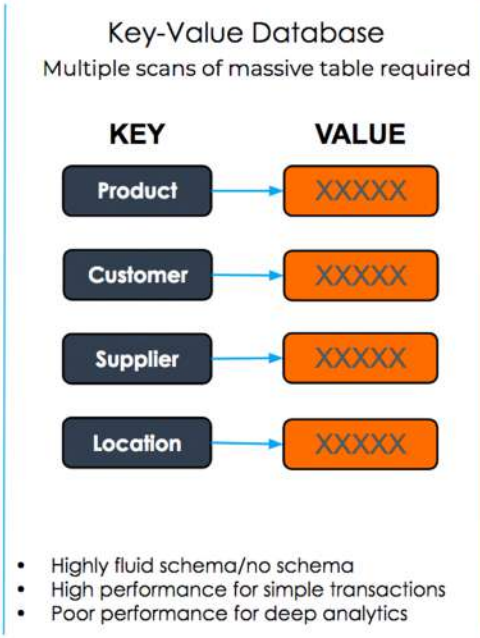
What is a Graph Database?

Graph analytics is the technology that allows for the deep exploration of complex interrelationships among various entities – organizations, people, transactions, pretty much any data that can be modeled and queried. Gartner estimates that the graph database and analytics market will grow at 100 percent annually through 2022, making it one of the fastest growing markets in data and analytics. According to Gartner, "Graph analytics will grow in the next few years due to the need to ask complex questions across complex data, which is not always practical or even possible at scale using SQL queries."

Relational and NoSQL Databases Aren't Up to the Task

Relational databases also fall short because their architecture simply isn't designed for this level of analytics. RDBs store the data for each business entity such as customer, order, product, and payment data in separate database tables. To understand and analyze relationships across the business entities, relational databases require table joins, which can take hours, even days for the complex joins and are computationally expensive as the size of the data grows. NoSQL databases store all of the data in a single table. This means that the relationship analysis requires scanning a huge table with millions or billions of rows, making it very difficult to perform a deeper analysis of the relationships beyond two or three levels.





Learn more in our Native Parallel Graphs eBook https://www.tigergraph.com/ebook and download the free TigerGraph Developer Edition at https://www.tigergraph.com/download

TigerGraph

What's a Graph Database and Why Should You Care?

Older Graph Technology Falls Short

Graph databases go back to the early 2000s and have long been seen as a great way to answer questions about complex relationships in large data sets. However, for a long time they struggled to perform well when data volumes grew large and when the answers must be provided in real time.

First-generation graph databases were built with native graph storage but were not made to handle large data or query volumes or perform beyond three levels or connections – known as hops -- inside the graph. With every hop in a graph, the scope of the search expands dramatically and the insights gleaned become deeper.

Second-generation graph databases were built on top of NoSQL storage, which allowed them to load large amounts of data. However, they still do not scale for queries involving three or more hops. Older graph databases also typically do not support "database sharding" – partitioning of data across a number of servers to increase scalability -- which means, a large graph with terabytes of data can't be distributed. These legacy graph databases are ill-equipped to scale up to today's real-world requirements, which call for a system that can perform many hops efficiently and in parallel to deliver sub-second query performance on big data.

The Evolution of Graph Databases	Graph 1.0 Single server, non-parallel	Graph 2.0 NoSQL base for storage scale	Graph 3.0 Native, Parallel
Native Graph Storage	✓	✗ Key-value or column store	✓
Parallel Loading	✗ Days to load terabytes	✗ Days to load terabytes	✓ Hours to load terabytes
Parallel Multi-Hop Analytics	✗ Times out after 2 hops	✗ Runs out of time/memory after 2 hops	✓ Sub-second across 10+ hops
Parallel Updates (in real-time)	✗ Batch updates	✗ Batch updates	✓ Mutable/ Transactional
Scale up and out for Speed and Size	✗ Single-server graph only	✓ Sharded graphs, handles 1-2 hop queries	✓ Sharded graphs, 2B+ queries/day for 10+ hops
Privacy for Sensitive Data	✗	✗	✓ MultiGraph

"We quickly ran into problems scaling with our original graph database – loading the data took a lot of time and once it was loaded, calculations either didn't finish or it was extremely slow."

- Noel Gomez, Data Sciences Leader, Amgen

Watch the testimonial at https://tinyurl.com/TGtestimonials

Leap Forward with a Native Parallel Graph Database - TigerGraph

TigerGraph is a new kind of graph database, a native parallel graph database purpose-built for loading massive amounts of data (terabytes) in hours and analyzing as many as 10 or more hops deep into relationships in real-time. TigerGraph supports transactional as well as analytical workloads, is ACID compliant, scales up and out with database sharding.

TigerGraph's proven technology supports applications such as fraud detection, customer 360, IoT, AI and machine learning to make sense of ever-changing big data, and is used by customers including Visa, Intuit, China Mobile, Wish and Zillow.

Learn more in our Native Parallel Graphs eBook https://www.tigergraph.com/ebook
and download the free TigerGraph Developer Edition at https://www.tigergraph.com/download