TigerGraph

Top Trends in Data Science and Engineering 2022:
Graph Databases + Graph Data Science

Jay Yu, Ph.D. - VP Product & Innovation

November 11, 2021

About Me

Jay Yu Ph.D.
VP of Product & Innovation
TigerGraph

Leading TigerGraph's product, solution and innovation teams.

Hands-on full-stack innovator, strategic thinker, leader and evangelist for new technology and product, with 25+ years of industry experience, including big tech company like Teradata and Intuit.

Prior to TigerGraph, served as a Distinguished Architect / Director at Intuit, with focus on enterprise data architecture and strategy, led 4 teams successfully adopted TigerGraph.

Hold 29 patents and received a PhD from University of Wisconsin - Madison's parallel database engine research group.

About TigerGraph

Our product: **the most scalable and performant graph analytics platform**

- MPP natvie graph database with speed **40-300x faster** than competition
- First and the only vendor completed **36TB** LDBC (SNB-SF30K BI) benchmark
- Bring AI/ML to your data via **In-database Machine Learning**
- Supports HTAP (**OLTP + OLAP**) for real-time and batch workloads
- SQL-like query language (**GSQL**) and delightful dev experience (**GraphStudio**)
- Available **on-premise** & on leading cloud platforms (**AWS, GCP, Azure**)

Our customers: enterprises at all sizes, including **Fortune 500 companies**, in

- financial services, healthcare, telecom, media, utilities and innovative startups in cybersecurity, ecommerce and retail …

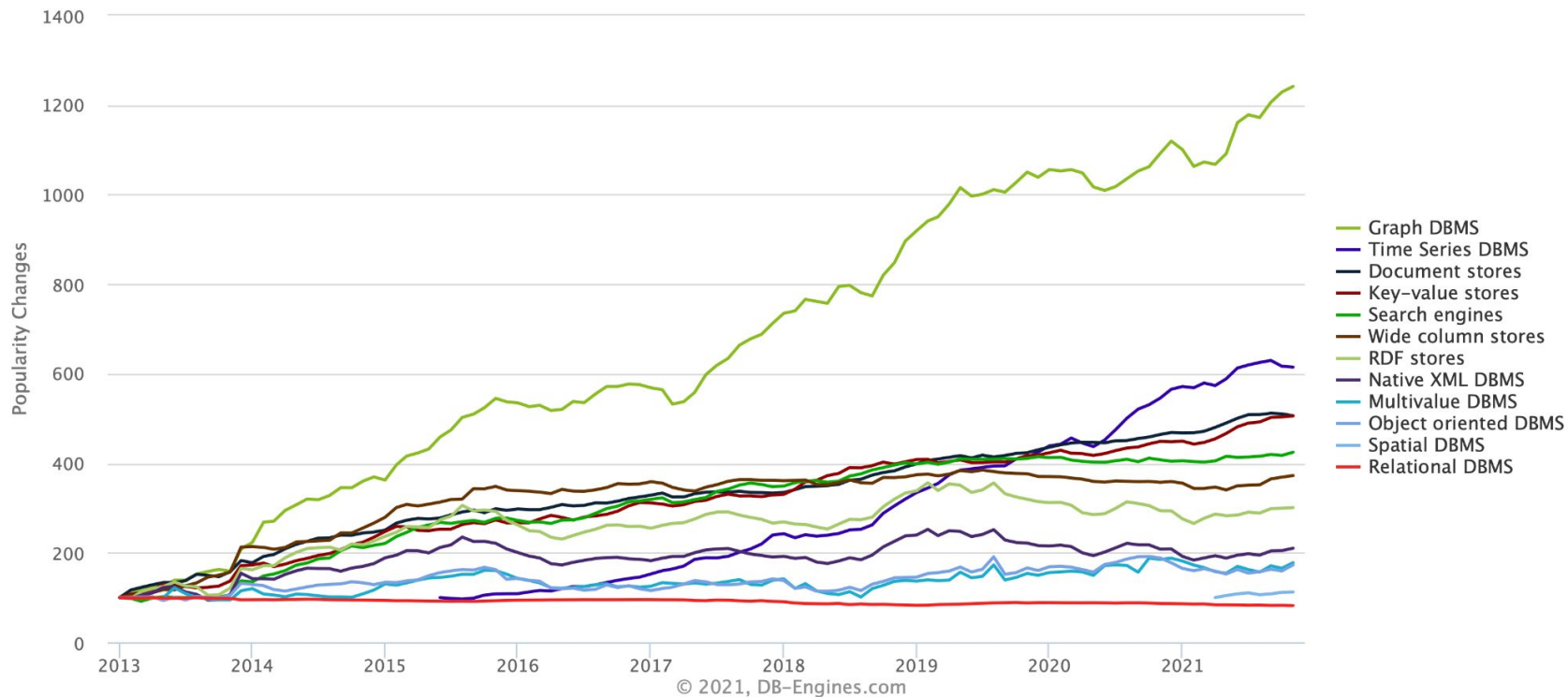Our company: **innovative and disruptive** database technology company

- Founded in 2012, HQ in Redwood City, California. **Total Funding: $170M+**.

2022 Top Trend in Data Engineering: Graph Databases

Rapid Growth of Graph Databases

Complete trend, starting with January 2013

Legend:
- Graph DBMS
- Time Series DBMS
- Document stores
- Key–value stores
- Search engines
- Wide column stores
- RDF stores
- Native XML DBMS
- Multivalue DBMS
- Object oriented DBMS
- Spatial DBMS
- Relational DBMS

© 2021, DB–Engines.com

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

By 2025, graph technologies will be used in 80% of data and analytics innovations, up from 10% in 2021, facilitating rapid decision making across the enterprise.

Gartner®

Source: Gartner, "Top Trends in Data and Analytics for 2021", Rita Sallam et al, 2021

Why Graph is becoming more and more Important ?

Data Silo

Volume of data created and replicated worldwide (source: IDC)

Zetabyte (ZB) Scale: 1,000,000,000 TB
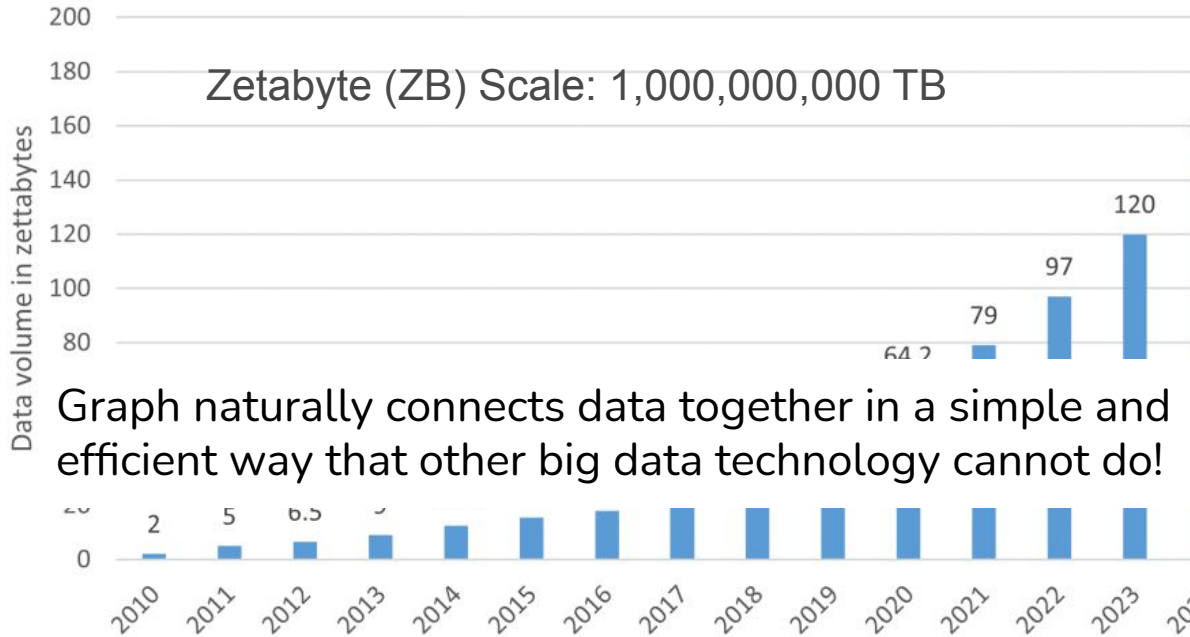
Graph naturally connects data together in a simple and efficient way that other big data technology cannot do!

Data volume in zettabytes

| 200 |
| 180 |
| 160 |
| 140 |
120	120
100	97
80	79
64.2

2 | 5 | 6.5

0

2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023

Data Swamp

TigerGraph

Graph Is How WE THINK

Graph is a natural, adaptable model for transforming data into knowledge and insights.

Locations

Events

Profile

Purchases

Browsing History

Transactions

Purchases

Locations

Profile

Events

Browsing History

8

Recommendations

New Products

Predictions

Supply Chain

CyberSecurity

Blockchain, Network Analysis

Identify key data and process massive amounts of data

Use the power of relationships and deep analytics to provide insights
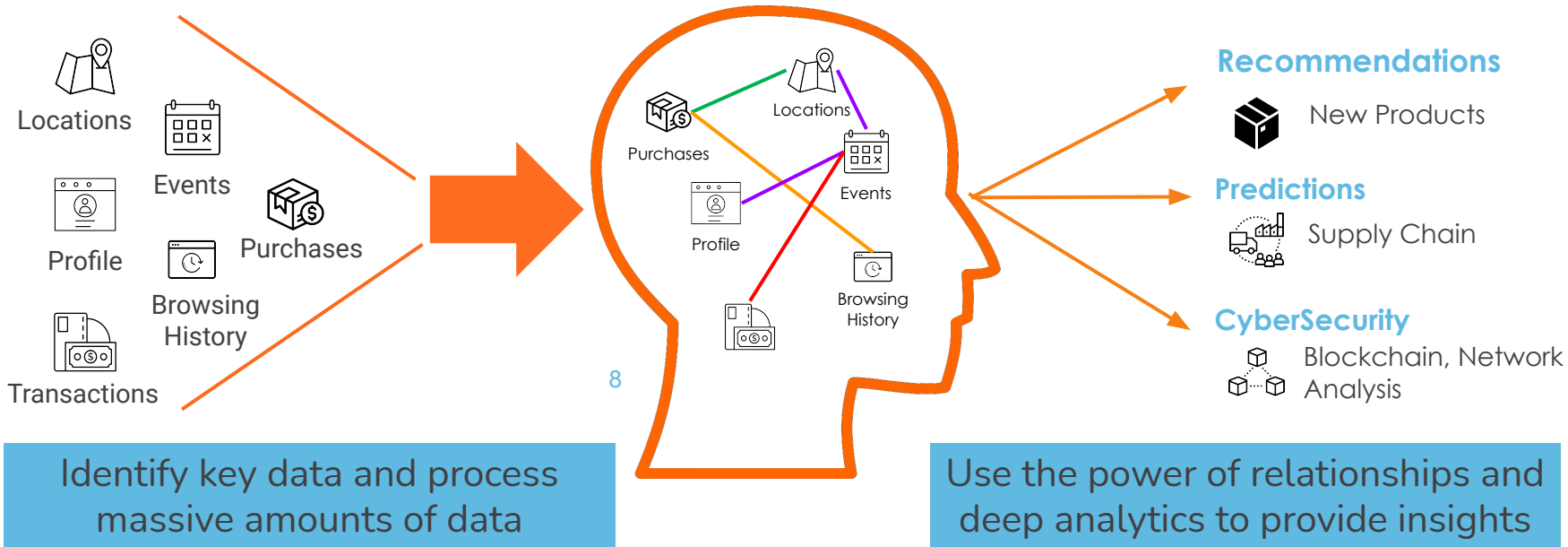
How is Graph Model different than other Data Models ?

Relational
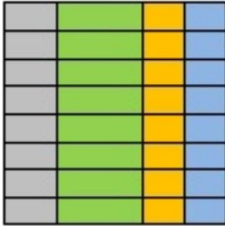
Analytical (OLAP)

Key-Value
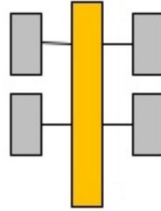
key → value
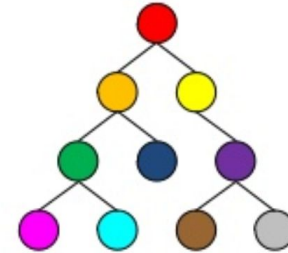key → value
key → value
key → value
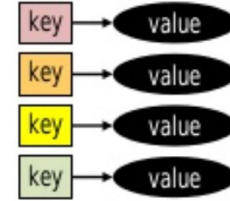
Column-Family
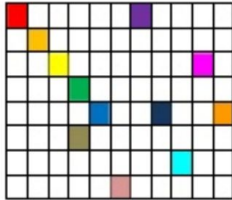
Graph
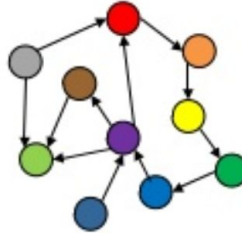
Document

Graph Model can represent all other Data Models **Naturally** !

Source: Dan McCreary https://www.slideshare.net/Dataversity/nosql-now-nosql-architecture-patterns-23589170

Adoption by Fortune 500 Enterprises

CONNECT ALL DATASETS AND PIPELINES

Friction-free scale up from GB to TB to Petabyte with lowest cost of ownership

UNITEDHEALTH GROUP®
Customer 360 connecting 200+ datasets and pipelines

Fortune 50 Retailer
Item 360 for eCommerce across 100+ datasets

xandr   AT&T
Identity graph connecting multiple data pipelines

Distributed Graph DB

TigerGraph

Advanced Analytics

In-Database Machine Learning
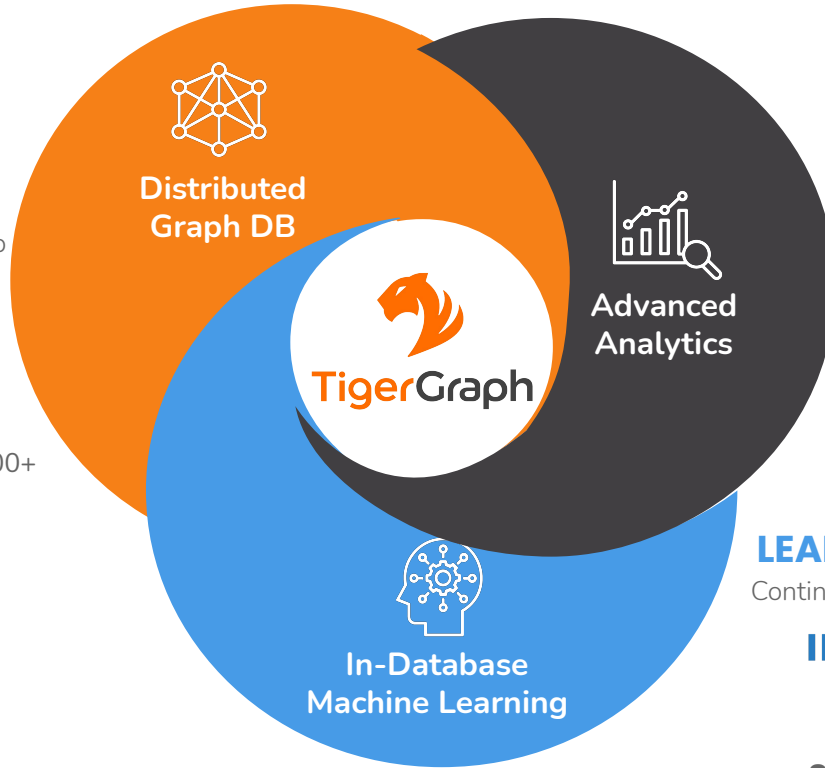
ANALYZE CONNECTED DATA
10-100X faster than current solutions

Jaguar Land Rover
Supply chain planning accelerated from 3 weeks to 45 minutes

China Mobile International
Fraud Detection - batch to real-time for 750 million calls/day

LEARN FROM CONNECTED DATA
Continuous graph-based feature generation and training

intuit
AI-based Customer 360 for entity resolution, recommendation engine, fraud detection

8 out of top 10 global banks
Real-time fraud detection and credit risk assessment

Industry Standard Benchmark Reaches New Scale !

Big Graph with 72.6B Vertices / 533.5B Edges / 36.1TB
Business Intelligence Queries w/ Incremental Batch Updates

Vertex Type	# of Vertices	Vertex Type	# of vertices
comment	58,666,958,815	company	1,575
post	13,148,296,221	university	6,380
forum	728,629,666	city	1,343
person	74,689,437	country	111
		continent	6
		tag	16,080
		tagClass	71
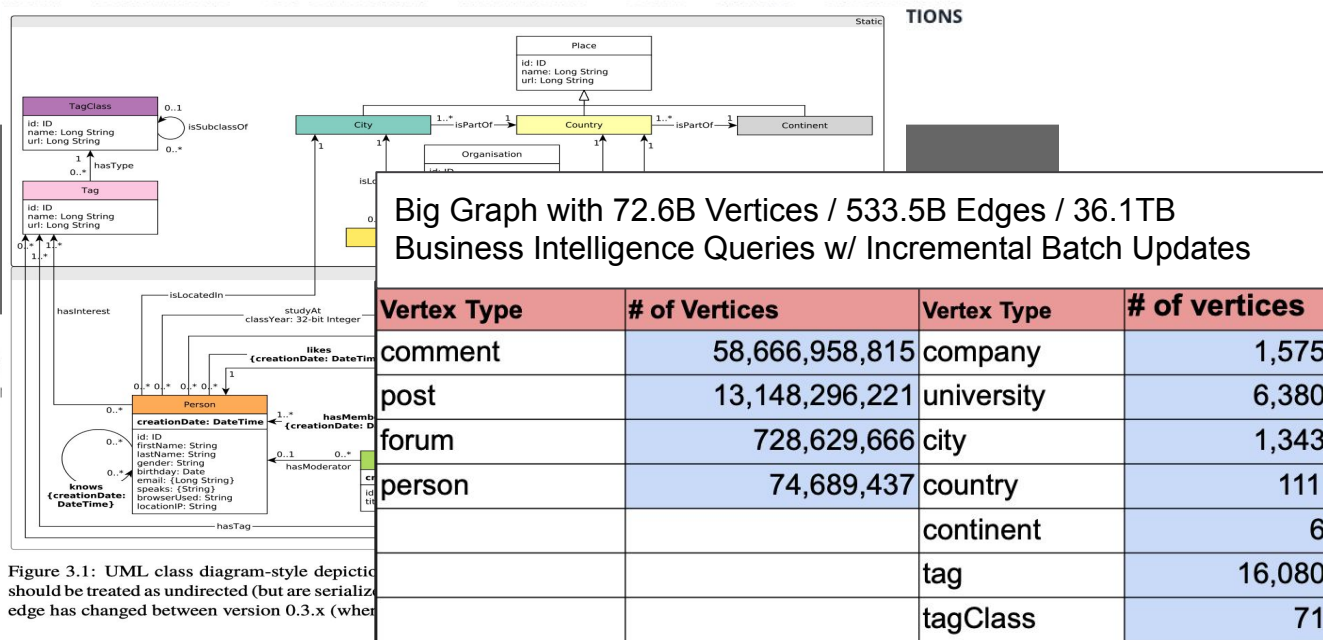
TigerGraph: *the first and only vendor* completed LDBC SNB-SF30K BI at 36TB!

Graph Query Languages are Converging: GQL !
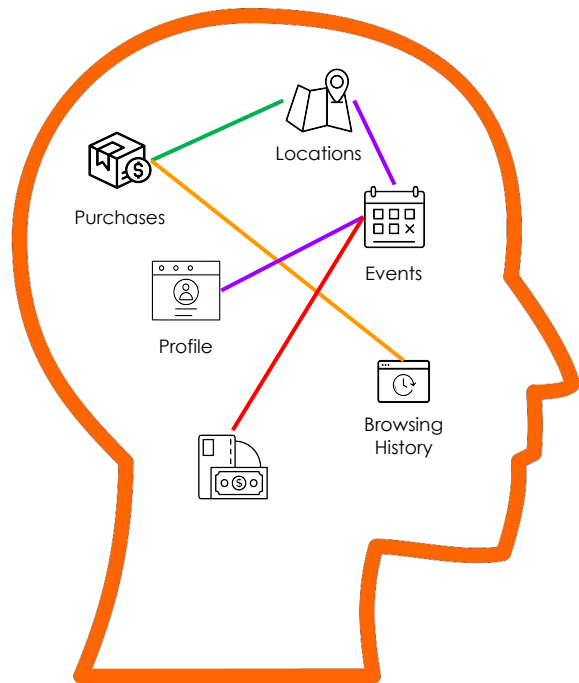

Source: https://www.gqlstandards.org/home
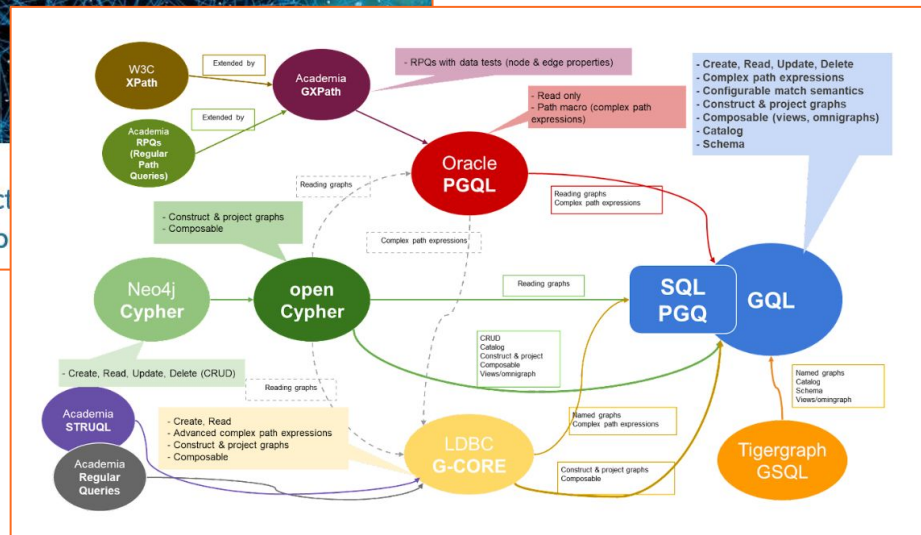
2022 Top Trend in Data Science: Graph Data Science

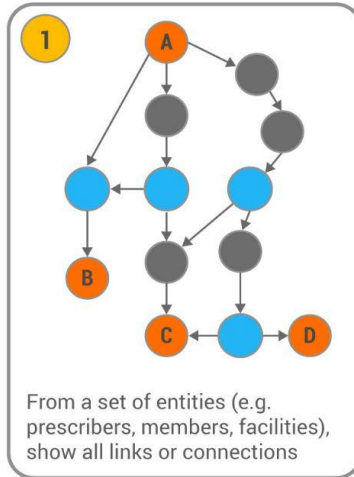# Rise of Graph Data Science

- Natural Merging of Data + Learning: Relationships are Fundamental

- Meaningful Patterns
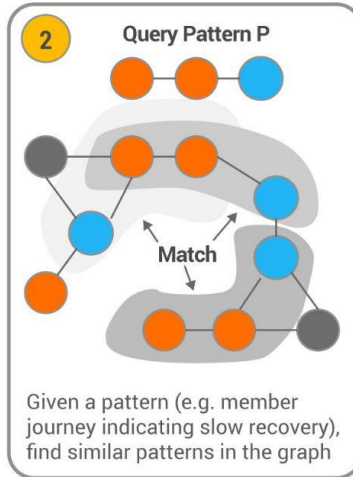
- Enhances Conventional ML

- Explainable Results

Graph Data Science brings Expert Knowledge to ML via Graph Analytics !
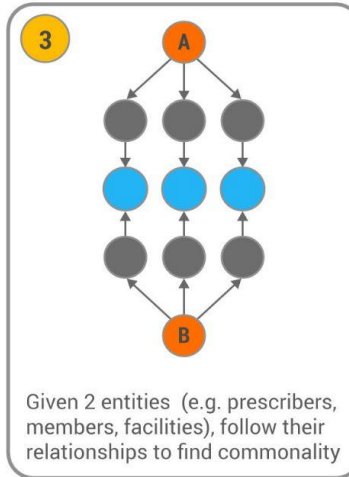
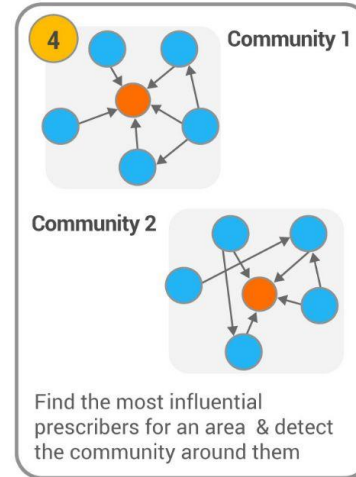# Graph Analytics: Unsupervised Learning

Deep Link Analysis

From a set of entities (e.g. prescribers, members, facilities), show all links or connections

Multi-dimensional Entity & Pattern Matching

Given a pattern (e.g. member journey indicating slow recovery), find similar patterns in the graph

Relational Commonality Discovery & Computation

Given 2 entities (e.g. prescribers, members, facilities), follow their relationships to find commonality

Hub & Community Detection

Find the most influential prescribers for an area & detect the community around them

---

5 Geospatial Graph Analysis

Analyze changes in entities (prescribers, members) & their relationships with location data

---

6 Temporal (Time-Series) Graph Analysis

Analyze changes in (prescribers, members) & their relationships over time

---

7 Machine Learning Feature Generation & Explainable AI

Extract graph-based features to feed as training data for machine learning; Power Explainable AI
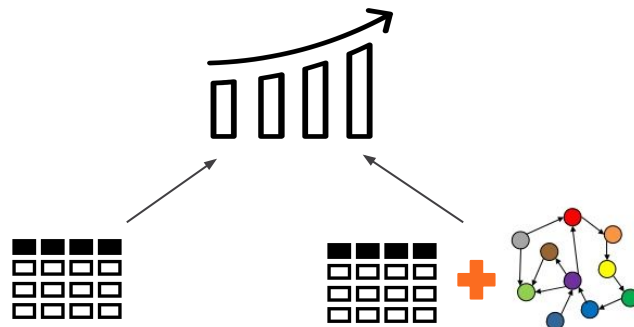
Graph Data Science by TigerGraph

In-Database Graph Algorithms

- 50+ algorithms for similarity, clustering, best paths, community

- No export needed. Live, updatable data

- Open-source, customizable

- Scaleable, Ultra-fast MPP graph engine

- Toolkit, building blocks for problem-solving

Graph Features Enriching ML Quality

- Graph enriches ML training data: quality in ⇒ quality out

- Use graph algorithms, pattern-matching queries, or graph embedding

Solving Enterprise-Scale Business Problems

Sample Graph Data Science Use Cases in Production

Healthcare:
Real-time recommendations

- 1.3TB graph brain
- Real-time care recommendations
- Improving healthcare, lowering cost

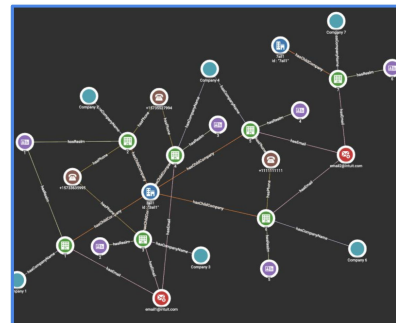## Industrial Supply Chain:
Analytics for decisions
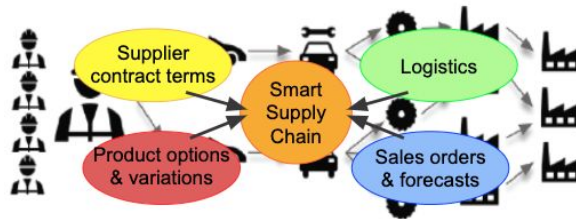
- Analytics: weeks→ minutes
- Reveal opportunities, optimize tactical & strategic decisions
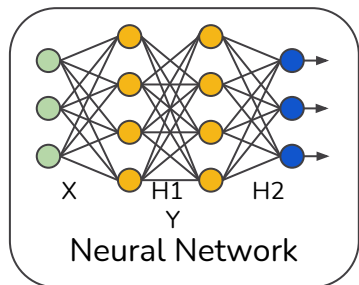- Saving $100M+/yr
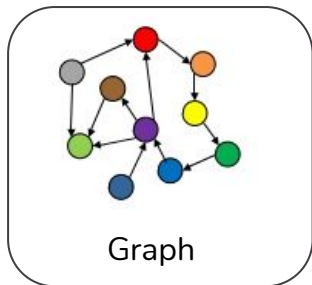
Financial Services:
Real-time fraud detection

- Integrates multiple tools
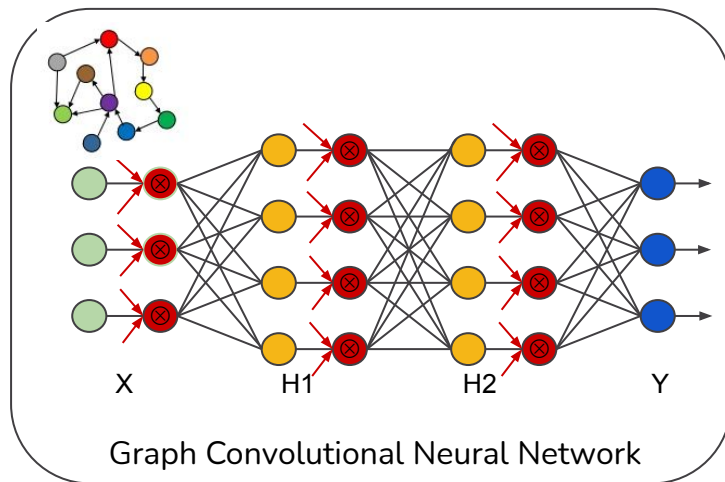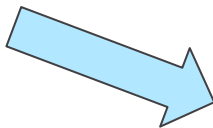- "Magical" real-time visual results for investigators
- Scalable for growth

Advanced R&D: Graph Neural Networks

Neural Network

Powerful Machine Learning to Predict and Classify

Graph

Insight from Connected Data
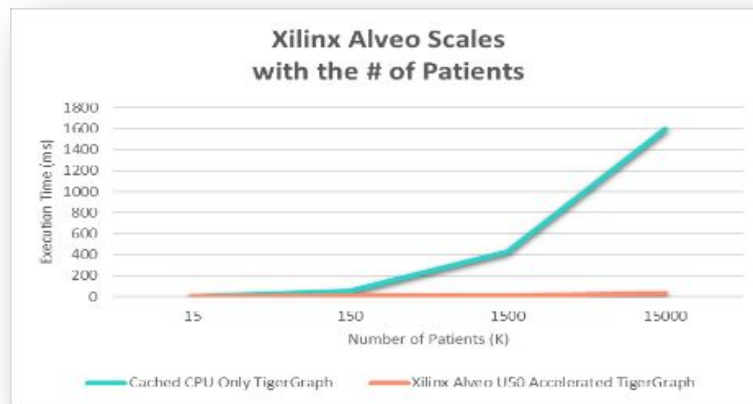
Graph Convolutional Neural Network

- Combines the added insight from connected data with the modeling power of neural networks

- Uses the graph during training; in-database training is the ideal.

HW Acceleration for Graph Data Science

Special HW Integration accelerating compute intensive graph algorithms by **20X to 48X !**

Summary

Key Takeaways: in 2022, we will see ...

The Year of Graph Databases !

- Scale: 10TB → 100TB

- Standardization: Siloed → GQL (Query) + LDBC (Benchmark)

- Deployment: On-Prem → Cloud (Private + Public)

Rapid Adoption of Graph Data Science !

- Setting: Exploration → Production

- Approach: ML Only → In-Database Machine Learning (Graph + ML)

- Infrastructure: Commodity HW → + Specialized HW Acceleration

![TigerGraph logo]

Try TigerGraph Cloud for free:

- tigergraph.com/cloud

Learn more from Graph + AI Summit

- tigergraph.com/graphaisummit

Thank You