How Intuit, Jaguar Land Rover, Xandr and UnitedHealth Group Are Driving Business Outcomes with Graph & AI

Gaurav Deshpande - VP of Marketing

It's Great to Meet You all Virtually

Gaurav Deshpande

Vice President of Marketing

- All things marketing at TigerGraph
- Led 2 startups through explosive growth - i2 Technologies (IPO) & Trigo Technologies (largest MDM acquisition by IBM)
- Patents in supply chain management and big data analytics
- **HUGE GRAPH HEAD**
- Email: gaurav@tigergraph.com

Who is TigerGraph?

We provide **advanced analytics and machine learning on connected data**

- The only scalable graph database for the enterprise: 40-300x faster than competition
- Foundational for AI and ML solutions
- Designed for efficient concurrent OLTP and OLAP workloads
- SQL-like query language (GSQL) accelerates time to solution
- Available on-premise & on: Google GCP, Microsoft Azure, aws

Our customers include:

- The largest companies in financial, healthcare, telecom, media, utilities and innovative startups in cybersecurity, ecommerce and retail

Founded in 2012, HQ in Redwood City, California

Corporate Overview Video

Advanced Analytics and Machine Learning on Connected Data

CONNECT ALL DATASETS AND PIPELINES

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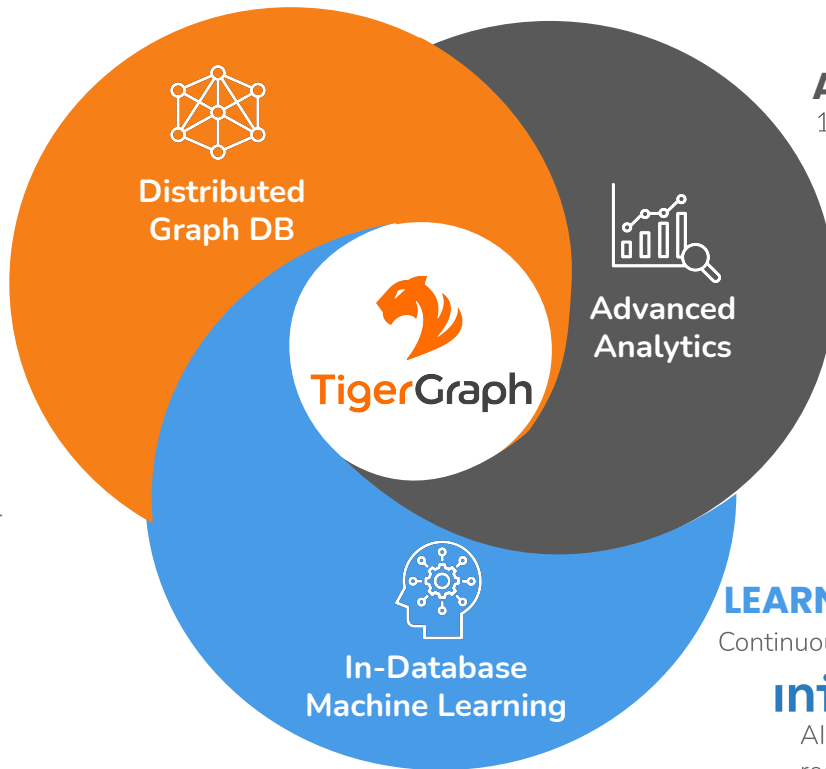
UNITEDHEALTH GROUP®

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

Advanced Analytics

TigerGraph

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

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

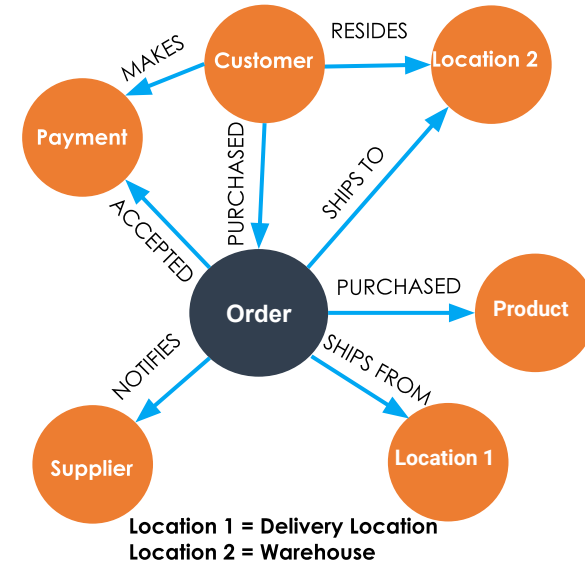# Evolution of The Database Landscape and The Rise of Graph DB

Relational Database

Product
xxxxxxx

Customer

xxxxxx

Order
xxxxxxxx

Supplier

xxxxxxx

Location
xxxxxxxx
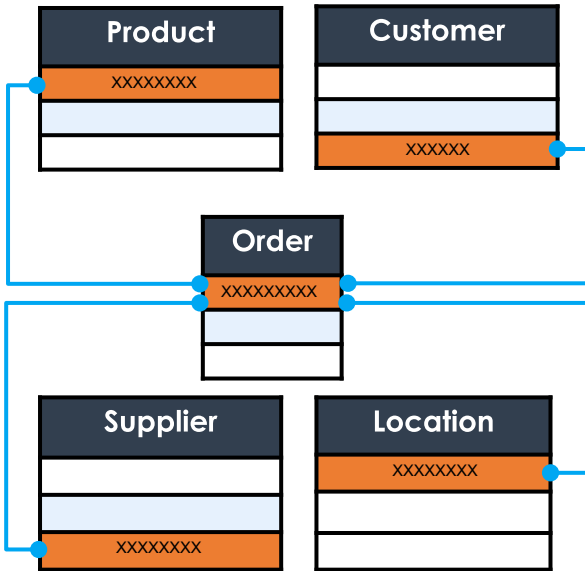
- Rigid schema
- High performance for transactions
- Poor performance for deep analytics

Key-Value Database

KEY   **VALUE**

Product → XXXXX

Customer → XXXXX

Supplier → XXXXX

Location → XXXXX
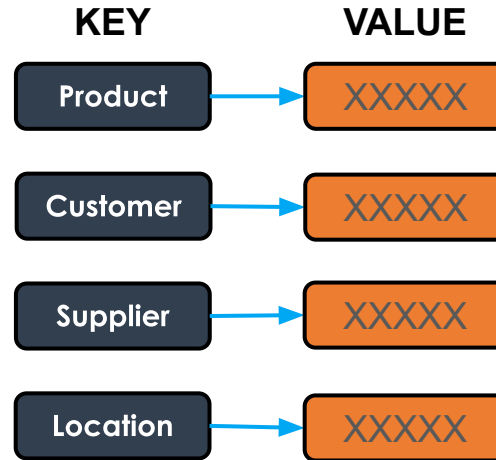
- Highly fluid schema/no schema
- High performance for simple transactions
- Poor performance deep analytics

Graph Database

MAKES   Customer   RESIDES
Payment   Location 2
PURCHASED   SHIPS TO
ACCEPTED   Order   PURCHASED → Product
NOTIFIES   SHIPS FROM
Supplier   Location 1

Location 1 = Delivery Location
Location 2 = Warehouse

- Flexible schema
- High performance for complex transactions
- High performance for deep analytics

Traditional Approach, with A Relational Database
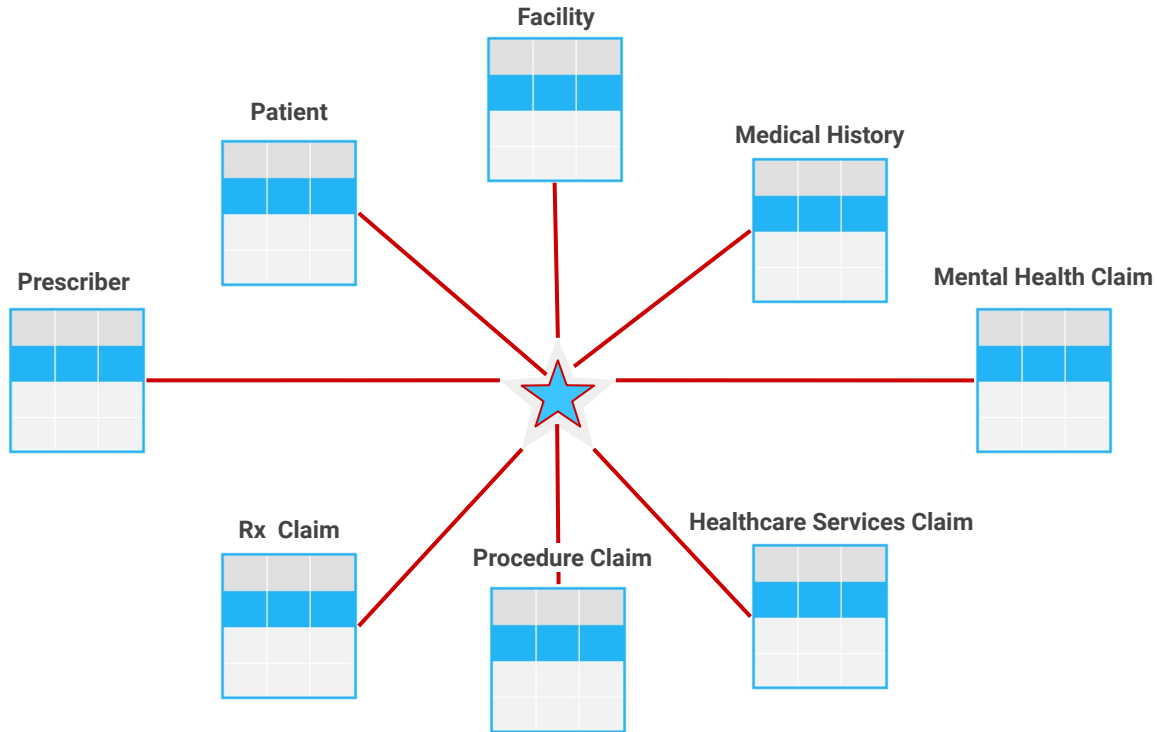
Complex Database Table Joins Across Silos Leads To Delayed Business Insights

Business Need 1: Connect Your Data

Distributed Graph DB

Advanced Analytics

In-Database Machine Learning

TigerGraph

CONNECT ALL DATASETS AND PIPELINES

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

UNITEDHEALTH GROUP®
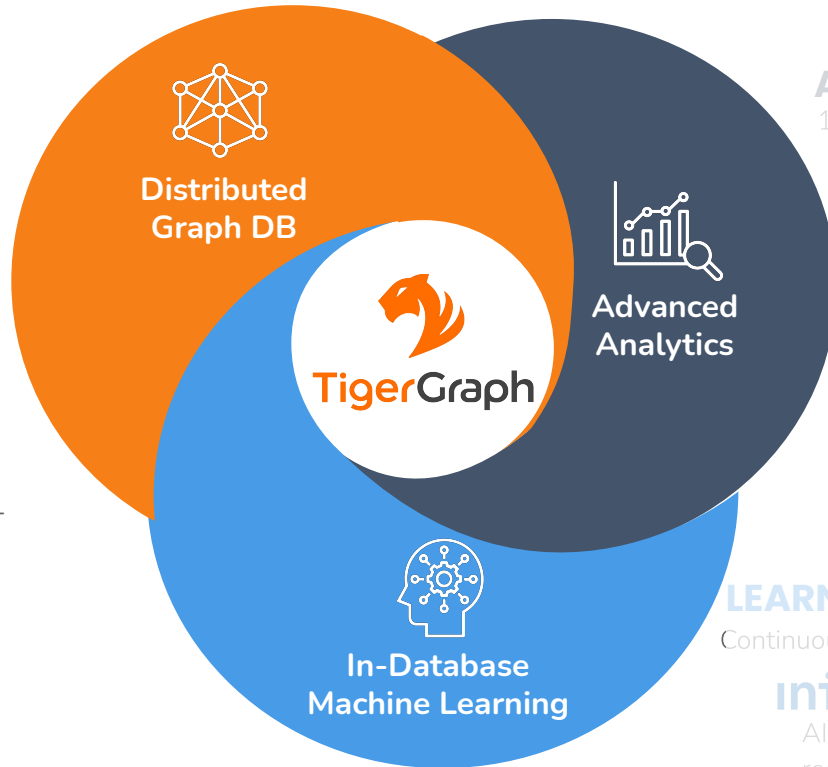Customer 360 connecting 200+ datasets and pipelines

Fortune 50 Retailer
Item 360 for eCommerce across 100+ datasets

xandr AT&T
Identity graph connecting multiple data pipelines

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 300 million calls/day

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

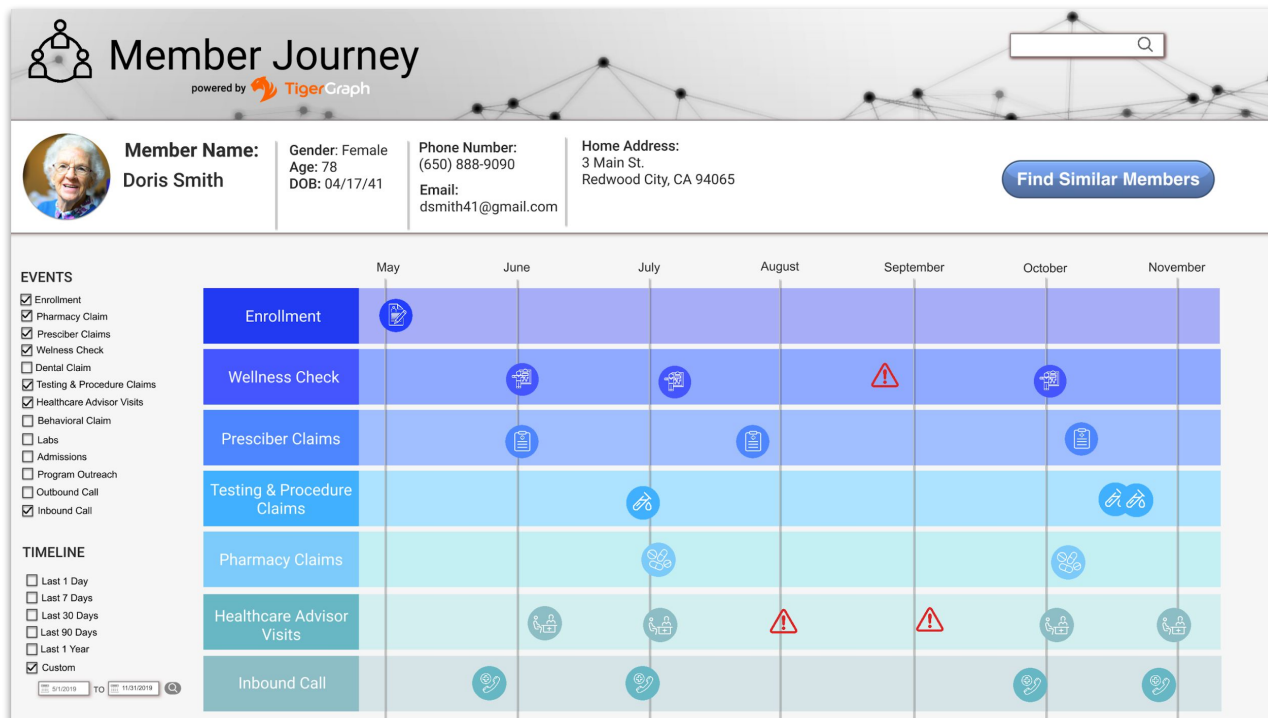**intuit**
AI-based Customer 360 for entity resolution, recommendation engine, fraud detection

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

Healthcare: **Connect All Datasets and Pipelines** for Customer 360 Journey for Healthcare Insurance Members

- **Integrate 200+ datasets and pipelines** to provide unified view for each member driving higher productivity for 23,000+ call center agents

- **Find similar members** with a click of a button in real-time (50 ms)

- **Deliver care path recommendations** for similar members

UnitedHealth Group Has Built the Largest Healthcare Graph in the world with 10 billion entities (claims, patients, doctors..), 50 billion relationships & 23,000+ users! (Graph+AI Keynote - https://info.tigergraph.com/keynote-edward-sverdlin)

⬡xandr

An Identity Graph stitches together different identifiers into a unified view of **people,** the **households** they belong to and **devices** they use.
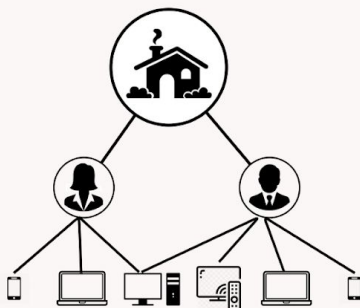
Why is Identity important?

People use multiple devices and screens daily

Identity enables cross-device & converged addressable advertising
* *More efficiency…Household/Consumer Frequency Capping*
* *More reach…Audience Extension to Linked Devices*
* *More lift…Conversion Attribution across Devices*

People expect relevancy and personalization

People demand privacy
Identity allows consent elections across device & affiliated brands.

People will be harder to reach & target in a 3rd party cookieless future
Deterministic 1st party ID consortiums of publishers & brands help.

From Graph + AI Conference Presentation by Xandr Data Science team - www.tigergraph.com/xandr

From Graph + AI Conference Presentation by Xandr Data Science team - www.tigergraph.com/xandr

Graph Model

A Broad Collection From Multiple Sources
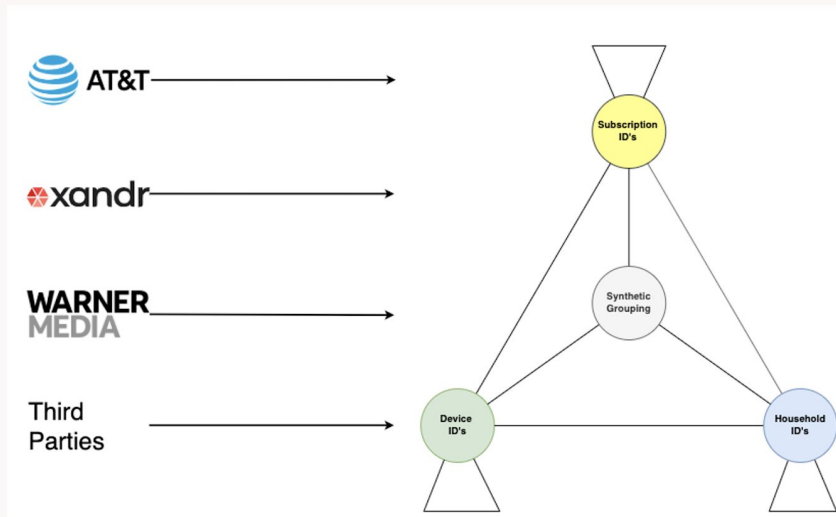
Subscription ID	Household ID
001	Aaa-001
001	Aaa-002
002	Bbb-001
003	Ccc-001
003	Ccc-002
003	Ccc-003

Household ID	Device ID	Subtype Property
Aaa-001	1002001	Cookie
Aaa-001	1002002	IDFA
Aaa-001	1002003	AAID
Aaa-001	1002004	Cookie
Bbb-001	1003001	Cookie
Bbb-001	1003002	AAID

From Graph + AI Conference Presentation by Xandr Data Science team - www.tigergraph.com/xandr

Does it scale?

- Distributed graph with 5+ billion vertices and 7+ billion edges
- Up to 1 billion daily graph updates from input
- 300 million vertices and 1+ billion edges created by the algorithms
- We built a 10 node TigerGraph cluster. Each node has 48 cores, 400GB RAM, 3GBps NVMe storage.
- Running BFS-style algorithms, like Label Persistence, spanning over a large distributed graph is extremely memory intensive
- We can add more RAM to the cluster nodes but vertical scaling has limits. We need to scale horizontally

From Graph + AI Conference Presentation by Xandr Data Science team - www.tigergraph.com/xandr

Xandr Improves Advertising Targeting Effectiveness with Identity Graph powered by TigerGraph

Opportunity

Xandr wanted a way to help advertisers target audiences with the right promotional messages by deeply analyzing data on consumers, devices, content, advertisers' needs and other attributes, collected across 15 WarnerMedia properties and credit reports from Experian.

Solution

Xandr has built an identity graph using AT&T, WarnerMedia, Third-party and its own data, and leverage TigerGraph to perform entity resolution.

Results

- Implement frequency-capping at the household or user level to ensure efficient advertiser spend
- Help advertisers find more consumers with audience extension and increase their campaigns lift with conversion attribution across different devices
- Manage consent elections across first party assets and third party data to respect customer preferences

More details including customer success story, Graph + AI conference session by Xandr team at **tigergraph.com/xandr**

Business Need 2: Analyze Connected Data

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

Advanced Analytics

In-Database Machine Learning

TigerGraph

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 300 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

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

Answering Critical Business Questions With Graph Analytics

Overview

Data and analytics leaders struggle to advance a shared understanding of data across business verticals and functions. Jaguar Land Rover demonstrates how graph analytics can give the business a connected view of supply and demand, enabling efficient answers to critical business questions.

Solution Highlights

1. Identify a common language for speaking business and data.
2. Connect supply and demand data in a knowledge graph and explore your most critical business problems by browsing up and down the graph. Examples:
   a) Demand for a model is suddenly surging in the US market. Do we have all the parts we need to meet this demand? Where do the supplier risks lie?
   b) Demand for a model is suddenly dropping drastically in the US market. What parts will we now have in surplus? How can we best use these parts?
   c) What is the profitability impact of changing a feature of a car?

About the Company

Jaguar Land Rover (JLR)

	Industry: Manufacturing Headquarters: Coventry, UK	Revenue: GBP 25.8 Billion (2019) Employees: 44,101 (2019)

Gartner case study for Jaguar Land Rover - Answering Critical Business Questions with Graph Analytics (Jaguar Land Rover), October 28, 2020, Gartner ID G00733557

Harry Powell
Director of Data and Analytics

Alice Grout-Smith
Data Scientist

Martin Brett
Senior Data Architect

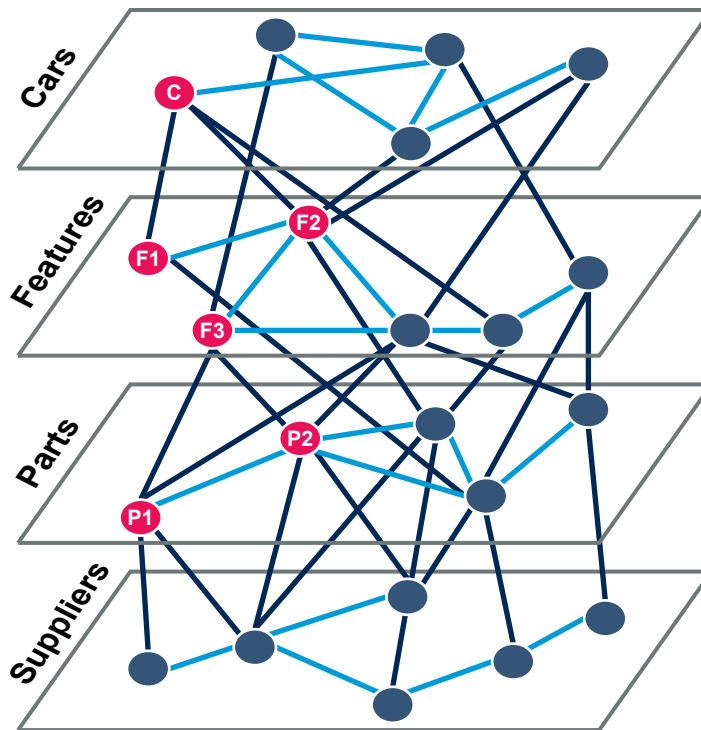**Hazel Scourfield**
Data Scientist

Clear Two-Way Line of Sight Between Demand and Supply

JLR's Demand-Supply Graph

Car C contains the feature F1.

Features F1 and F2 are connected because they are both features of car model C.

Parts P1 and P2 are connected because they are both parts for feature F3.

Demand

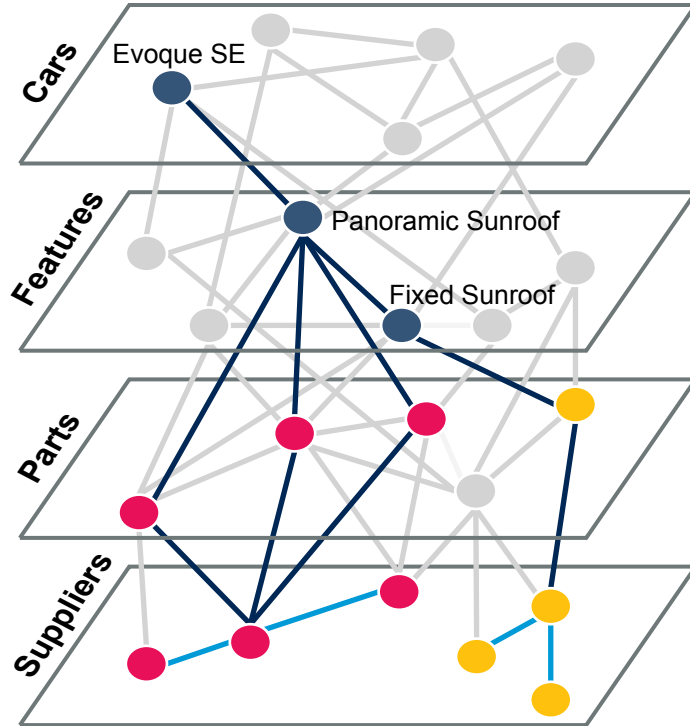When two features share a part or a car, the features are connected.

When two parts share a feature or a supplier, the two parts are connected.

Supply

Cars

Features

Parts

Suppliers

Identifying and Reducing Supply Chain Risks

JLR's Demand-Supply Graph for Exploration and Discovery

Cars — Evoque SE

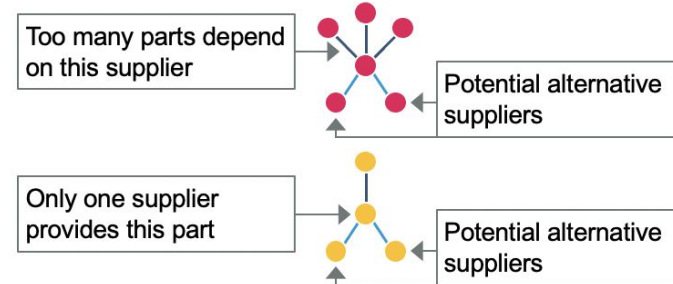Features — Panoramic Sunroof, Fixed Sunroof

Parts

Suppliers

Source: Adapted From Jaguar Land Rover

Critical Business Questions: Demand for the Evoque model is suddenly surging in the US market. Do we have all the parts we need to meet this demand? Where do the supplier risks lie?

Procedure for Exploration:

1. Browse the graph to identify all features for Evoque SE.

2. For each feature, browse the graph to find all parts needed for the feature.

3. For each part, browse the graph to identify all the suppliers for the part.

4. Look for graph substructures such as fan-in and fan-out patterns to identify and mitigate supplier risk.

Too many parts depend on this supplier → Potential alternative suppliers

Only one supplier provides this part → Potential alternative suppliers

Making the Most of Surplus Inventory

JLR's Demand-Supply Graph for Investigation and Inference
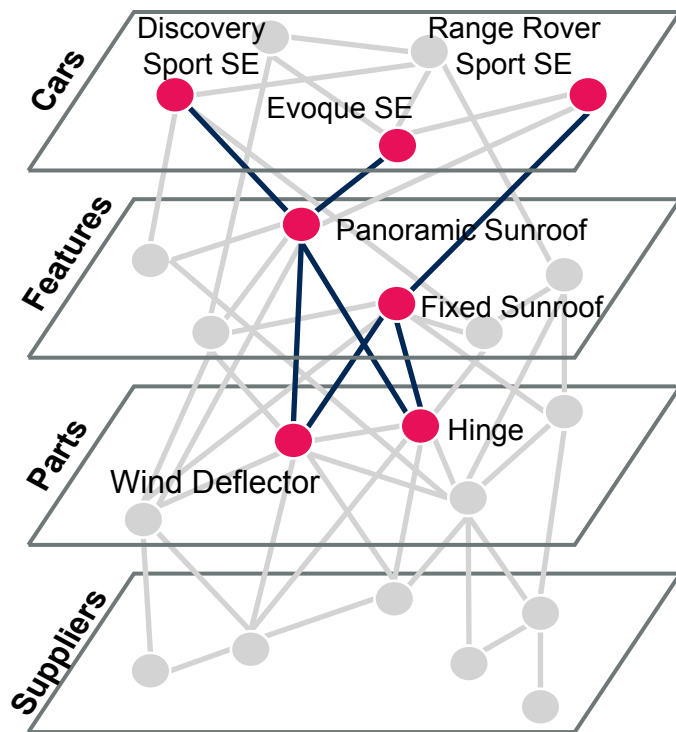
Source: Adapted From Jaguar Land Rover

Critical Business Questions: Demand for Discovery Sports is suddenly dropping drastically in the US market. What parts will we now have in surplus? How can we best use these parts?

Procedure for Investigation:

1. Browse the graph to identify all features of Discovery Sport SE. Let's call these features SF1.

2. Identify models that are in greater demand. For each model, identify the features to build the car. Let's call these features SF2.

3. Find the intersection of sets SF1 and SF2. These are the features this car model has in common with the Discovery Sport SE model. These features can be used to build this more in-demand car model instead.

4. For the features that are not shared between the car models, identify the parts that are used by the Discovery Sport SE alone. For each of these parts, look for the following fan-in and fan-out patterns.

Browsing Direction

This part could be used to build a feature for Range Rover Sport SE
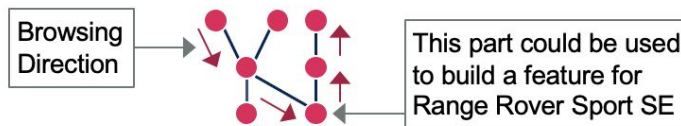
Solving an Intractable Optimization Problem

Critical Business Question: What is the profitability impact of changing a feature of a car?

Evoque With Plain Roof

((●)) One example of millions of what-if perturbations →
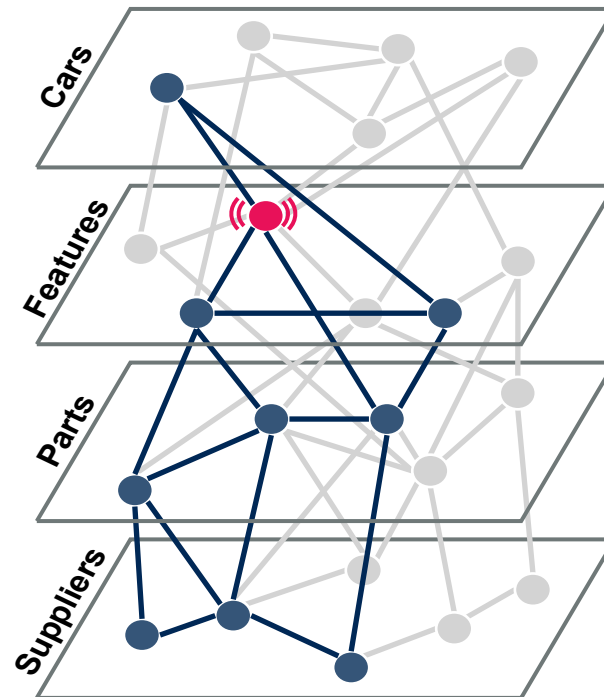
Evoque With Sunroof
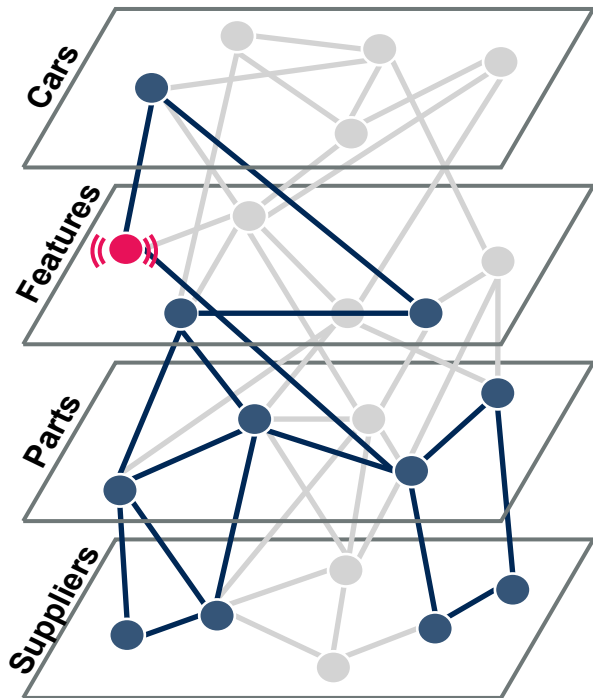
The feature change has upstream ripple effects on the car's price and **revenue**.

▲ **Revenue impact**

Replace the sunroof with the moonroof.

▼ **Cost impact**

The feature change has downstream ripple effects on parts inventory and **cost**.
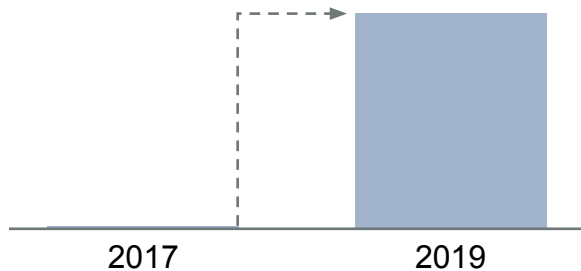
Source: Adapted From Jaguar Land Rover

Results

Decision Speed

▲ **120x**

2017   2019

Source: Adapted From Jaguar Land Rover

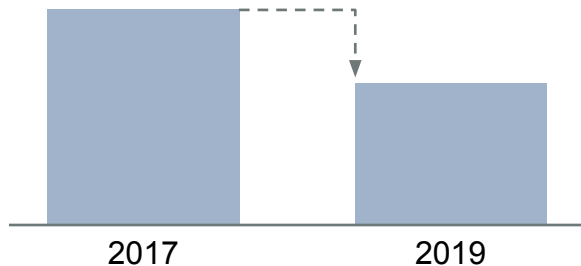## Business Value

▲ **3x**

2017   2019
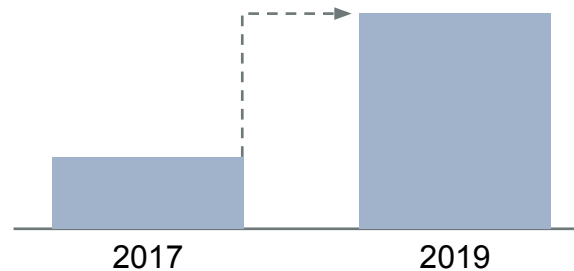
Source: Adapted From Jaguar Land Rover

Supplier Risk

▼ **35%**

2017   2019

Source: Adapted From Jaguar Land Rover

"As we began using the same data as our commercial and manufacturing partners, it has become a lot easier to work together and address our business problems in greater depth."

Director of Purchasing, JLR

Jaguar Land Rover Featured on CIO.com

Harry Powell
Director of Data & Analytics, JLR

Accelerate planning at JLR - weeks to minutes

The software, from TigerGraph, detected when suppliers would fail to meet quota demands.

"We used the graph to re-sequence how our vehicle orders were to be built in our factory in response to a supplier failure," Powell says. **Queries across the supply chain model now take 30 to 45 minutes compared to weeks using SQL relational database software.**

CIO.com article - **The pandemic pivot: IT leaders innovate on the fly, August 13 2020**

More details including customer success story, Graph + AI conference session at *https://www.tigergraph.com/customers*

Business Need 3: LEARN from Connected Data

CONNECT ALL DATASETS AND PIPELINES

Friction-free scale up from GB to TB to Petabyte with lowest cost of ownership
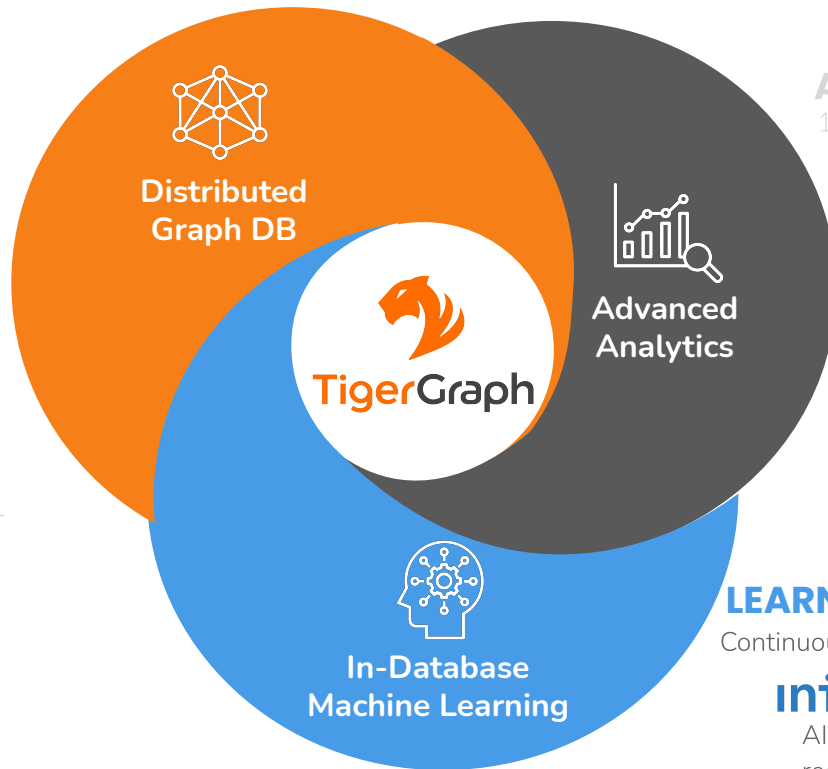
UNITEDHEALTH GROUP
Customer 360 connecting 200+ datasets and pipelines

Fortune 50 Retailer
Item 360 for eCommerce across 100+ datasets

xandr **AT&T**
Identity graph connecting multiple data pipelines

Distributed Graph DB

Advanced Analytics

TigerGraph

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 300 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

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

Fraud Detection: Need for Better AI

$30B Lost to Fraud

80% false positives (blocking non-fraud transactions)

- $118 billion of blocked sales in the U.S. with 15% of cardholders experiencing blocked sales

- High-income consumers (> $75,00/yr) at higher risk of false positives (22%)

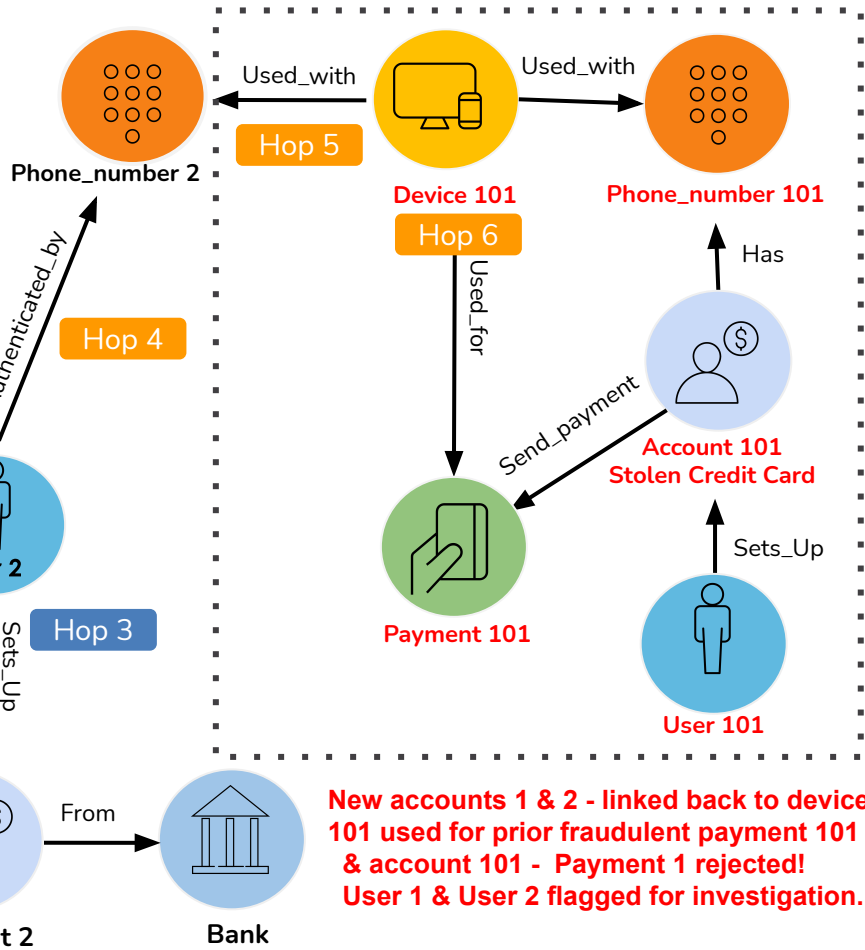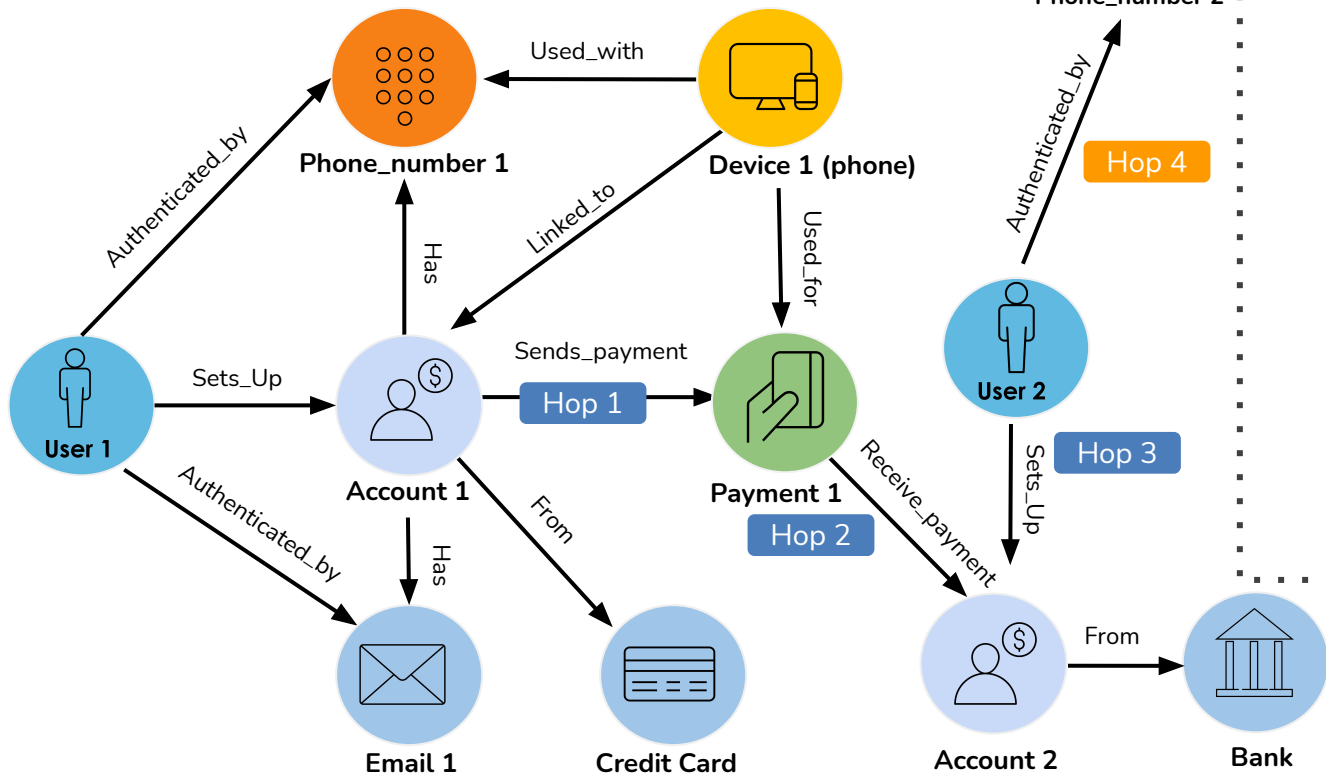- 40% of denied users are attempting to pay a greater than $250 transaction

Fraud Detection in Financial Services

Regular Analytics (Shallow) VS **Advanced Analytics(Deep) with TigerGraph**
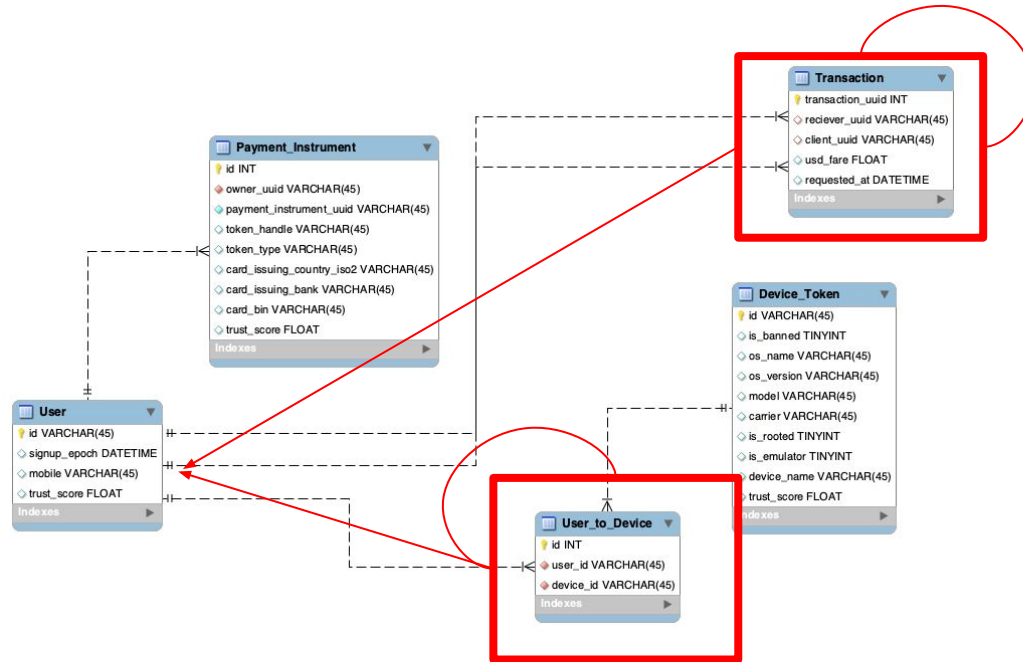

New accounts 1 & 2 - linked back to device 101 used for prior fraudulent payment 101 & account 101 - Payment 1 rejected! User 1 & User 2 flagged for investigation.

Sign up FREE for TigerGraph Cloud to use the starter kit for fraud detection (payments)

RDBMS Requires Complex Table Joins: Can't Support Real-Time Traversal of Connected Data

Preventing Fraudulent Loans with TigerGraph Tier 1 U.S. Bank

Business Challenge

A leading U.S. bank needed to search across **20TB** of data for possible connections between credit card applications known to be fraudulent and applications of unknown status - relational databases and other graph providers were not up to the task, as they were unable to deliver the **speed and scale** required.

Solution

- Pairing **graph technology with machine learning** to identify fraudulent activity at scale and intervene **in real-time**.

- Leveraging deep analytics to find hidden connections across 20TB+ of data.

Business Benefits

- Able to score and prevent fraudulent loan applications on a massive scale – minimum **30% uplift** and **$15M annual incremental fraud avoidance**. **$1.5M through cost savings** on false positives.

20TB
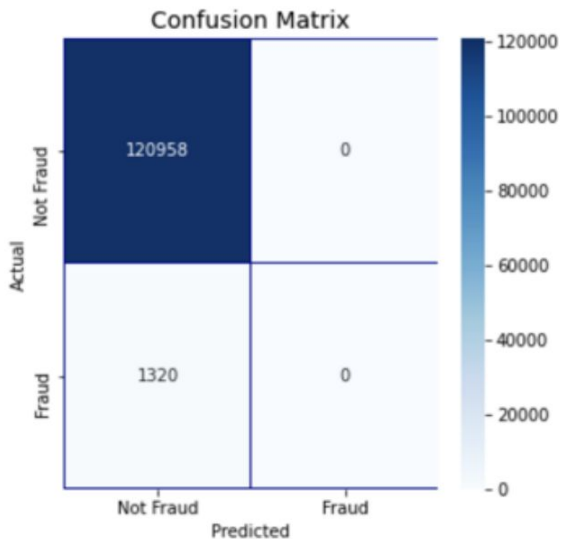Card applications data

6 weeks
PoC elapsed time

3 months
Time to build and fully deploy platform to production

$16.5M
1st year ROI with 30% uplift in fraud detection

CLV Impact > $100M

Feature Extraction for ML/AI

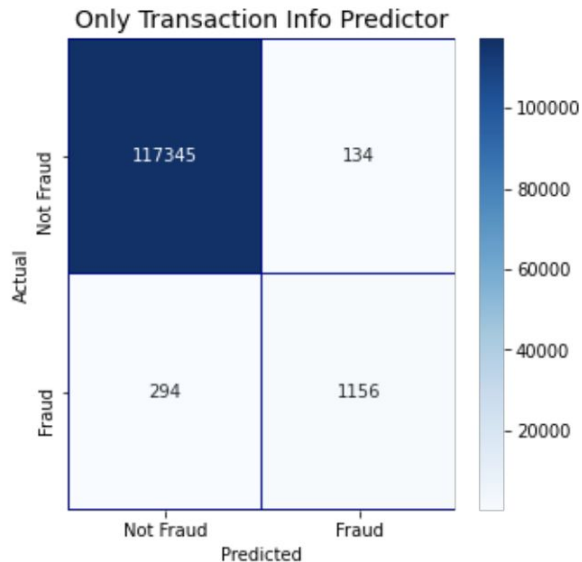| Variable description	Variables in use currently	Variables usable with a graph database
Total value of all transactions for merchant		1
Total number of frauds detected for merchant		2
Total value of all transactions for merchant category		3
Total number of frauds detected for customer		4
Total number of transactions performed by customer		5
Amount of transaction	3	6
Maximum value of one transaction for merchant		7
Customer	4	8
Total number of frauds detected for merchant category		9
Maximum value of one transaction for merchant category		10
Merchant category	2	11
Merchant	1	12
Total number of transactions for merchant category		13
Total number of transactions for merchant		14
Total amount of all transactions by this customer		15
Maximum value of a transaction for customer		16
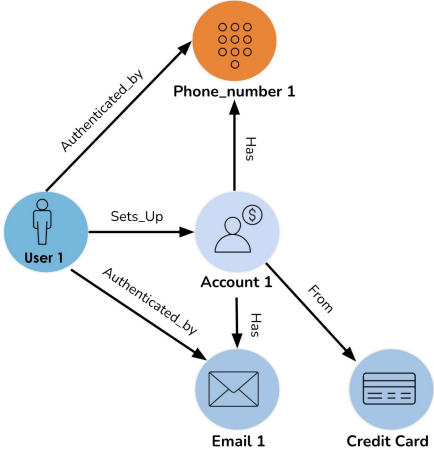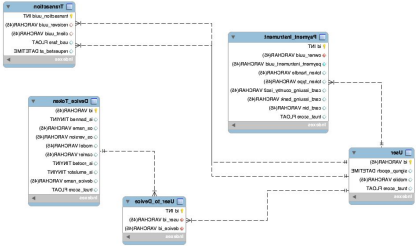
Double the performance of Fraud Detection System with 50% reduction in false positives & 50% reduction in undetected fraud transactions with Graph DB features when compared to legacy solution

Improvement

Confusion Matrix

	Not Fraud (Predicted)	Fraud (Predicted)
Not Fraud (Actual)	120958	0
Fraud (Actual)	1320	0

Only Transaction Info Predictor

	Not Fraud (Predicted)	Fraud (Predicted)
Not Fraud (Actual)	117345	134
Fraud (Actual)	294	1156

With Graph Predictor

	Not Fraud (Predicted)	Fraud (Predicted)
Not Fraud (Actual)	117424	55
Fraud (Actual)	153	1297

No Fraud Detection:

1320 frauds undetected
0 false positives

Best Classifier on Transaction Data:

294 frauds undetected
134 false positives
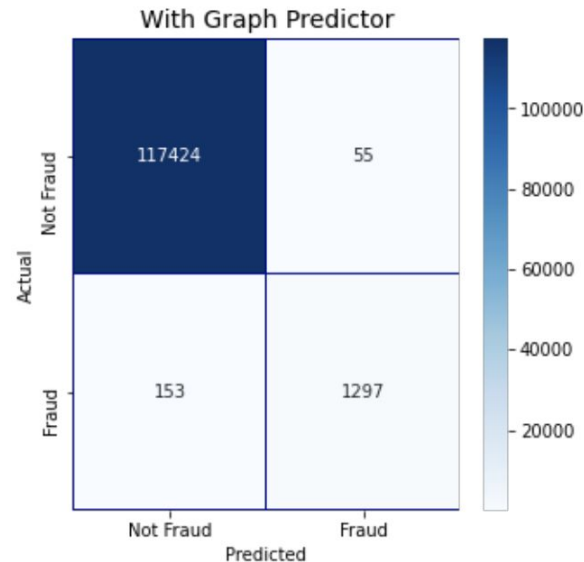
Best Classifier on Graph Data:

153 frauds undetected: 48% better
55 false positives: 59% better

Visit tigergraph.com → Solutions → Financial Services for the solution brief & machine learning workshop for building the fraud detection system with TigerGraph

Detecting Fraud Rings with TigerGraph Tier 1 U.S. Bank

Business Challenge

A leading U.S. bank wanted a better way to detect and remove fraudsters from their credit-card network. Prototypes had shown that a combination of advanced graph algorithms gave significant gains – big-data tools and other graph technologies either couldn't scale to the full customer base or gave inconsistent results.

Solution

- Implementing PageRank and Louvain [fraud] community detection in an MPP native-parallel database.

- Leveraging deep analytics to find hidden connections across 20TB+ of data.
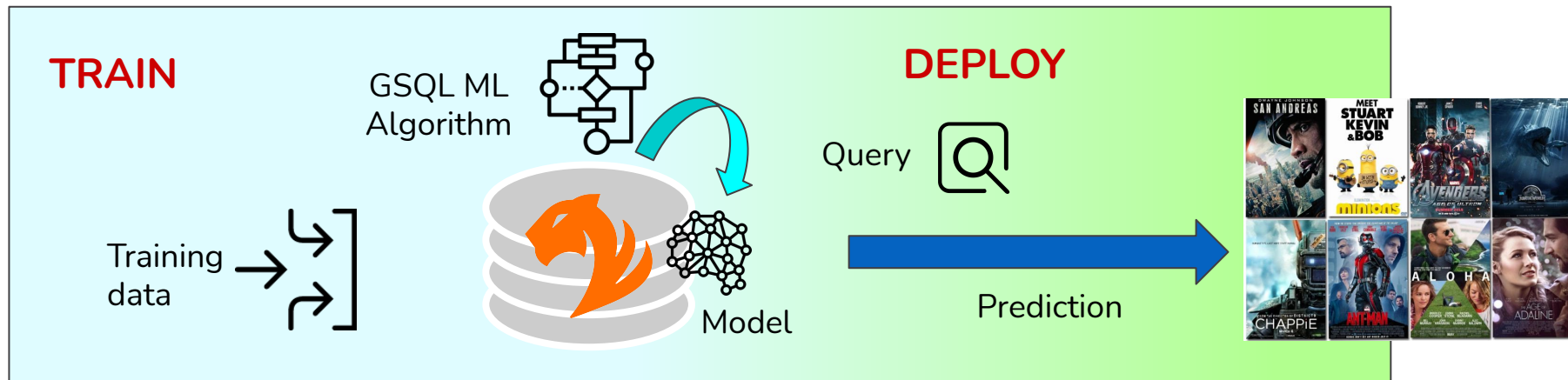
Business Benefits

- Able to expose fraud rings, shut down connected cards, and combat fraudulent activity on a massive scale –**35% uplift** and **$50M incremental fraud avoidance**. >**$1.5 million through cost savings** on false positives, infrastructure and TCO

10TB
Card applications data

6 weeks
PoC elapsed time

3 months
Time to build and fully deploy platform to production

+$50M
1st year ROI with 35% uplift in fraud detection

CLV Impact > $200M

In-Graph Database ML with GSQL

TRAIN

GSQL ML Algorithm

Training data

Model

DEPLOY

Query

Prediction

- Native graph storage and PG model
- Coded once, auto scale-out & scale-up
- Real-time updates
- GSQL Turing-complete language
    - Preprocess data
    - Training: flow-control, accumulator, pattern match
    - Model validation

Applications:

- Entity resolution
- Recommendation
- Fraud detection
- ...

In-Graph Machine Learning

Efficient Machine Learning requires:

	TigerGraph	typical graph database
Big Data Capacity	✔	?
Programmable for iterative, data-intensive algorithms	✔	✗
Parallel processing	✔	✗
Parallel accumulation	✔	✗

7 of the Top 10 Global Banks Use TigerGraph

Credit Card Fraud: Is applicant connected to potential fraudsters?

Impact Analysis: Communities or Clusters impacted by the fraud rings

Trade Surveillance: Are employees following the rules?

Merchant Analytics: Transaction sequencing to detect geolocation proximity.

Credit Scoring: Real-time credit scoring to help recommend offers best suited to customer profiles?

Wealth Management: What Accounts, HNI to target for stocks or life change events.

Semantics in Entity Resolution: Graph-based Entity Mastering
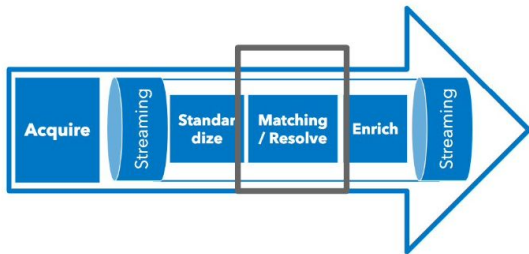
Near-real-time Incremental Entity Matching
- Semantics and lineage preserved
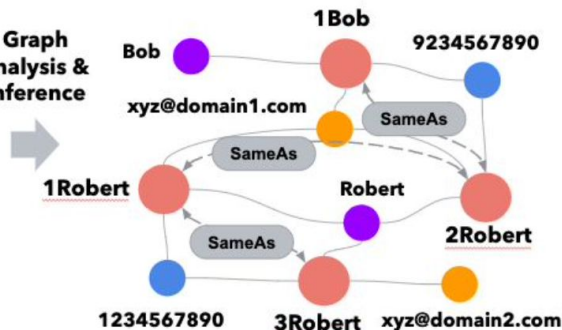- Declarative logic (inference, derivation)
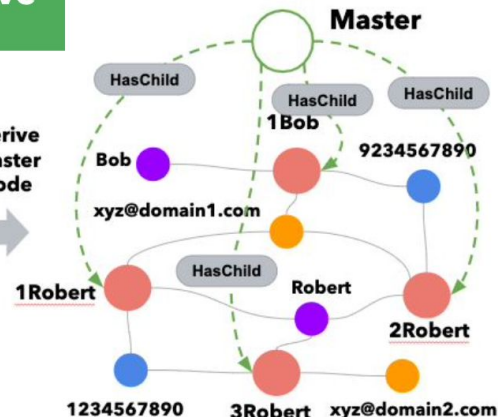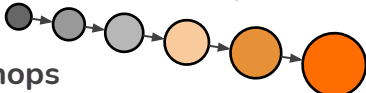- Built-in visual explanation
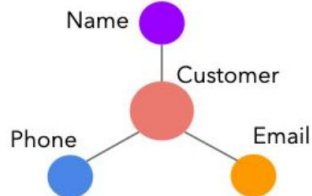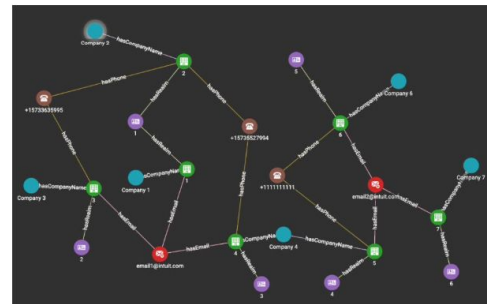
19% Reduction
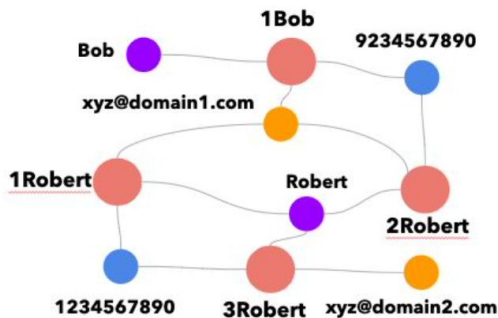
<1% False positive

The TigerGraph Difference

Feature	Design Difference	Benefit
Real-Time Deep-Link Querying **5 to 10+ hops**	• Native Graph design • C++ engine for high performance • Storage Architecture	• Uncovers hard-to-find patterns • Operational, real-time • HTAP: Transactions+Analytics
Handling Massive Scale	• Distributed DB architecture • Massively parallel processing • Compressed storage reduces footprint and messaging	• Integrates all your data • Automatic partitioning • Elastic scaling of resource usage
In-Database Analytics & Machine Learning	• GSQL: High-level yet Turing-complete language • User-extensible graph algorithm library, runs in-DB • ACID (OLTP) & Accumulators (OLAP)	• Avoids transferring data • Richer graph context • Graph-based feature extraction for supervised machine learning • In-DB machine learning training
NO CODE	• No-code migration from RDBMS • No-code Visual Query Builder	• Democratize self-service analytics to derive new-insights from legacy/external data stores

Starter Kits and Developer Portal for Graph+ML

TigerGraph CLOUD

1. Content-based movie recommendation: *similarity*, k-nearest neighbor + latent factor

2. Entity resolution: Link & merge similar entities, based on *similar* properties and neighbors

3. Low-rank approximation of graph relationships

4. Graph feature engineering for anti-fraud ML

Select a Starter Kit *

Pick a Starter Kit with sample graph data schema, dataset, and queries (e.g. Fraud Detection, Recommendation Engine, Supply Chain Analysis, etc.).

Additional information including overview video at **tigergraph.com/starterkits**

All   Anti-fraud

Geospatial Analysis

Graph Algorithms   Healthcare

Knowledge Graph   **Machine Learning**

Recommendations

In-Database Machine Learning Recommendation

In-Database Machine Learning for Big Data Entity Resolution

Low-Rank Approximation Machine Learning

Machine Learning and Real-time Fraud Detection

Developer Portal

dev.tigergraph.com

1. Unsupervised Learning with Graph Algorithms

2. Feature Set Extraction for Machine Learning

3. ML Enrichment with Graph Features

4. Graph Enrichment with Machine Learning

5. In-database ML Techniques for Graphs

Get Started for Free

- Get the Free Enterprise License at tigergraph.com

- Try TigerGraph Cloud with free tier - tigergraph.com/cloud

- Learn from 40+ on-demand sessions at tigergraph.com/graphaiworld

- Take a Test Drive - Online Demo at tigergraph.com/testdrive

- Join the Community at tigergraph.com/community

@TigerGraphDB /tigergraph /TigerGraphDB /company/TigerGraph