



Knowledge Graph and Machine Learning: 3 Key Business Needs, One Platform

Speaker: Dr. Victor Lee, Director of Product Strategy
and Developer Relations - TigerGraph

Hosted by: Kurt Cagle,
Community Editor - Data Science Central

December 16, 2020



It's Great to Meet You all Virtually

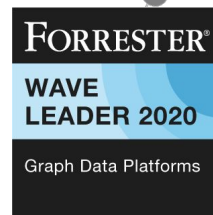


Dr. Victor Lee

Head of Product Strategy & Developer Relations, TigerGraph

- BS in Electrical Engineering and Computer Science from UC Berkeley, MS in Electrical Engineering from Stanford University
- PhD in Computer Science from Kent State University focused on graph data mining
- 20+ years in tech industry
- **Email: victor@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 the competition
- Foundational for AI and ML solutions
- Efficient concurrent OLTP and OLAP workloads
- SQL-like query language (GSQL) accelerates time to solution
- On-premises & on  Google GCP,  Microsoft Azure, 



Our customers include:

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



Founded in 2012, HQ in Redwood City, California

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®

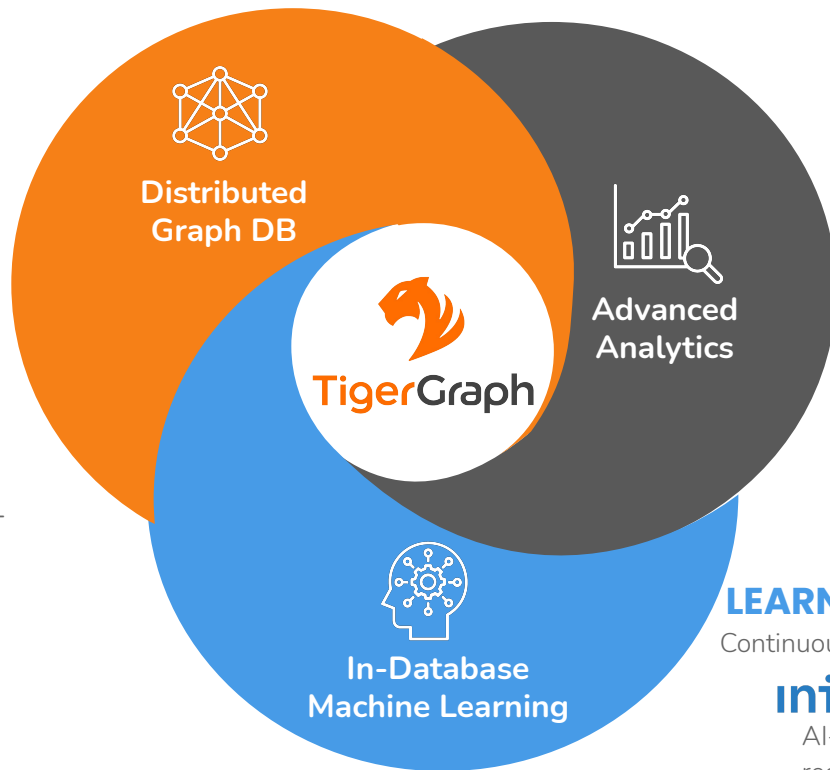
Customer 360 connecting 200+ datasets and pipelines

Fortune 50 Retailer

Item 360 for eCommerce across 100+ datasets



Identity graph connecting multiple data pipelines



ANALYZE CONNECTED DATA

10-100X faster than current solutions

Jaguar Land Rover

Supply chain planning - 3 weeks to 45 minutes



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

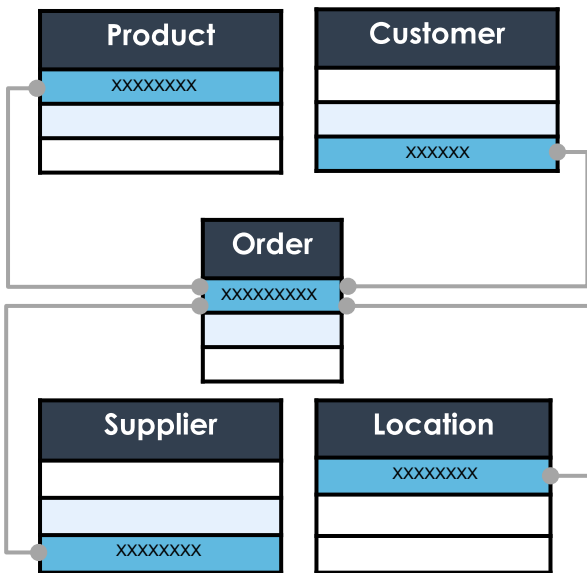
7 out of top 10 global banks

Real-time fraud detection and credit risk assessment



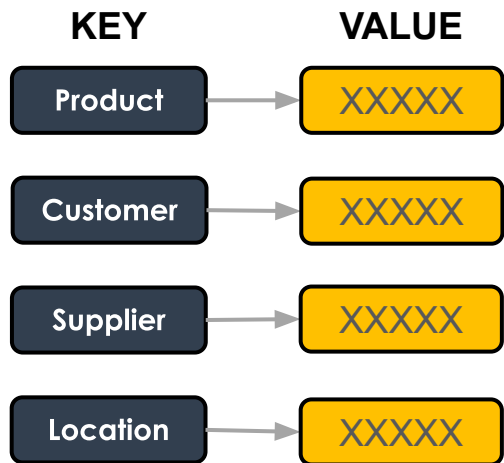
Evolution of The Database Landscape, The Rise of Graph DB

Relational Database



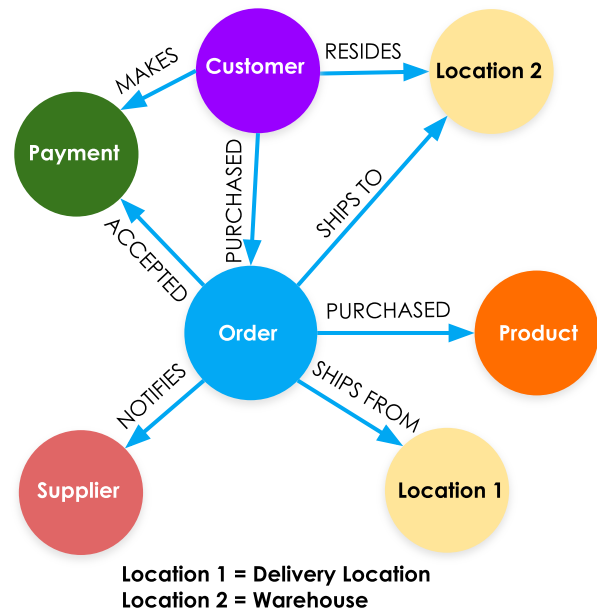
- Rigid schema
- High performance for transactions
- Poor performance for deep analytics

Key-Value Database



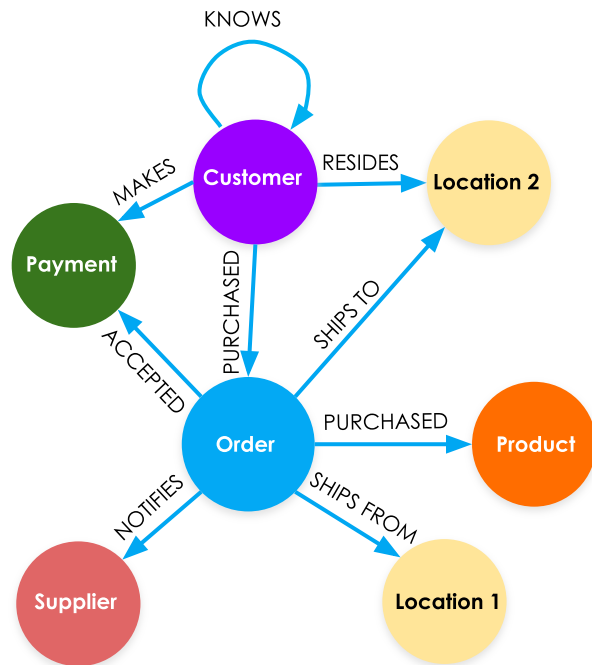
- Highly fluid schema/no schema
- High performance for simple transactions
- Poor performance deep analytics

Graph Database



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

Graph+AI Delivers More Value, Better Results



Richer, Smarter Data

- Connections-as-data
- Connects different datasets, breaks down silos

Deeper, Smarter Questions

- Look for semantic patterns of relationship
- Search far & wide more easily & faster than other DBs

More Computational Options

- Graph algorithms
- Graph-enhanced machine learning

Explainable Results

- Semantic data model, queries, and answers
- Visual exploration and results

Business Need 1: **Connect** Your Data

CONNECT ALL DATASETS AND PIPELINES

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

UNITEDHEALTH GROUP®

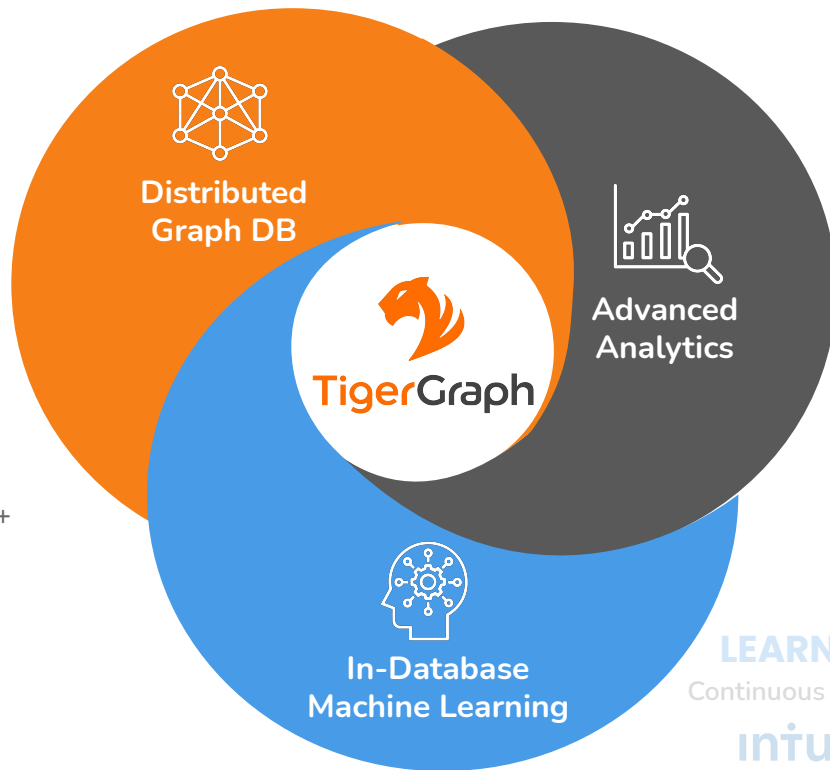
Customer 360 connecting 200+ datasets and pipelines

Fortune 50 Retailer

Item 360 for eCommerce across 100+ datasets



Identity graph connecting multiple data pipelines



ANALYZE CONNECTED DATA

10-100X faster than current solutions

Jaguar Land Rover

Supply chain planning - 3 weeks to 45 minutes



Fraud Detection - batch to real-time for 300 million calls/day

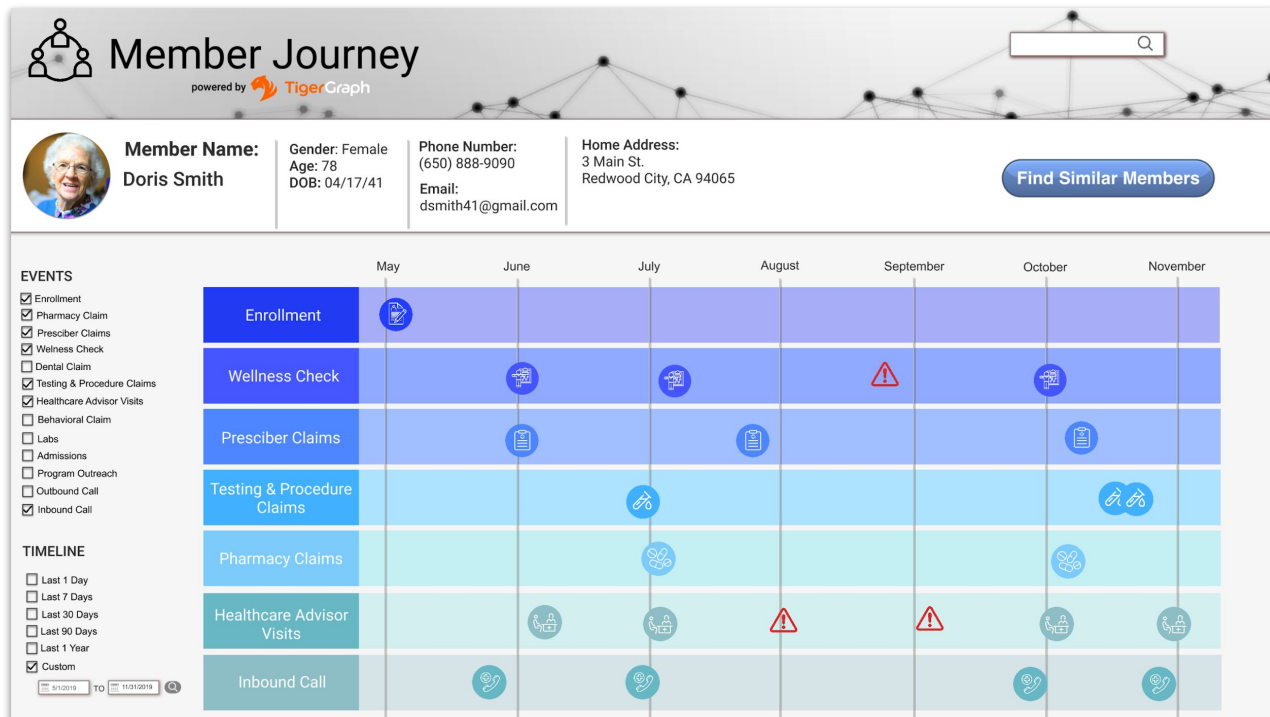
LEARN FROM CONNECTED DATA

Continuous graph-based feature generation and training

intuit.

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

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 call center operations
- **Find similar members** with a click of a button in real-time (50 ms)
- **Deliver care path recommendations** for similar members

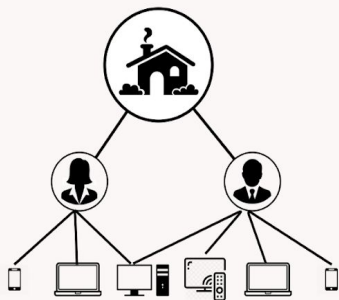
UnitedHealth Group Has Built the Largest Healthcare Graph in the world with 10 billion entities (claims, patients, doctors..), 50 billion relationships & 23,000+ users! (Graph+AI Keynote - <https://info.tigergraph.com/keynote-edward-sverdlin>)

Media/Telecom: Entity Resolution of people & households

What is an identity graph?



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

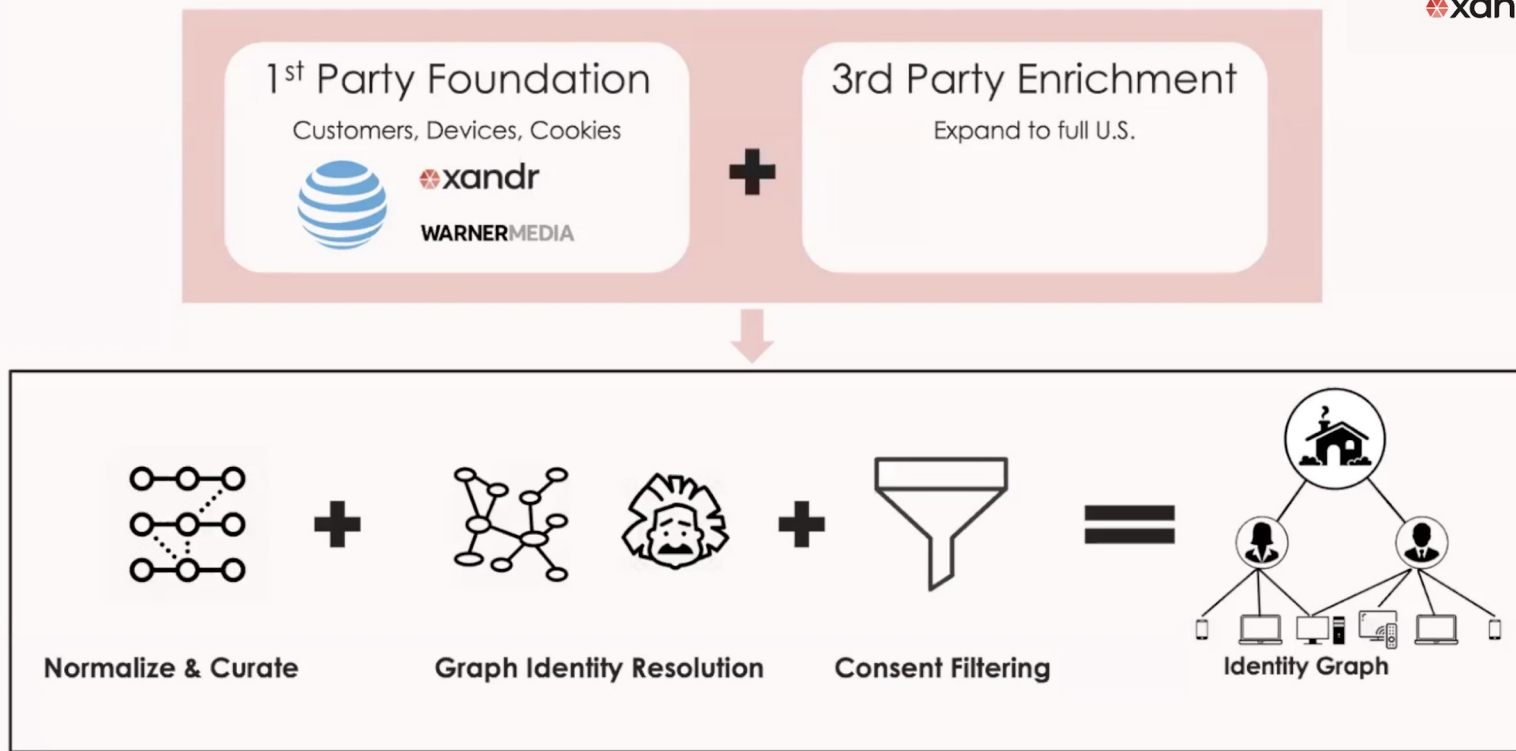
Identity allows consent elections across device & affiliated brands.

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



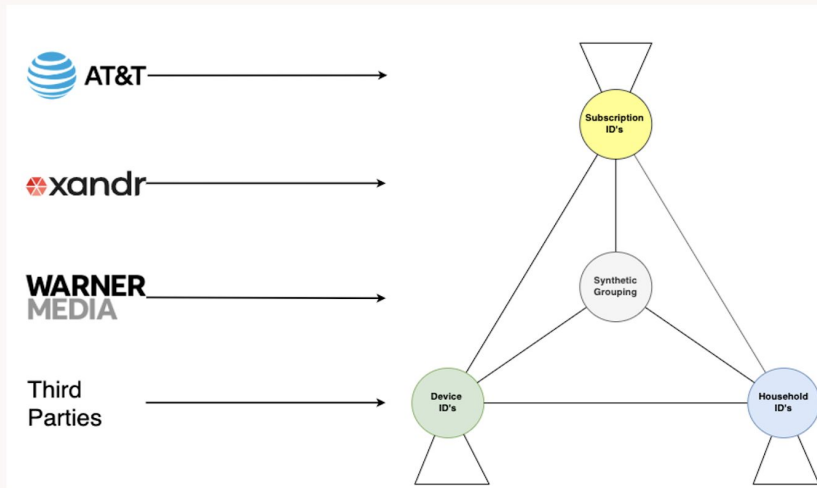


From Graph + AI Conference Presentation by Xandr Data Science team - www.tigergraph.com/xandr

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



- Distributed graph with 5+ billion vertices and 7+ billion edges
- Up to 1 billion **daily** graph external **updates**
- 300 million **vertices** and 1+ billion **edges created by the algorithms**
- 10-node TigerGraph cluster. Each node has 48 cores, 400GB RAM, 3BGps NVMe storage
- BFS-style **algorithms [in GSQL]**, like Label Persistence, over a large distributed graph
- We can add more RAM [scale vertical]. We need to **scale horizontally [add server nodes]**

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 deterministic and probabilistic 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: customer success story, Graph+AI conference session by Xandr team www.tigergraph.com/xandr



Business Need 2: **Analyze** Connected Data

CONNECT ALL DATASETS AND PIPELINES

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

UNITEDHEALTH GROUP*

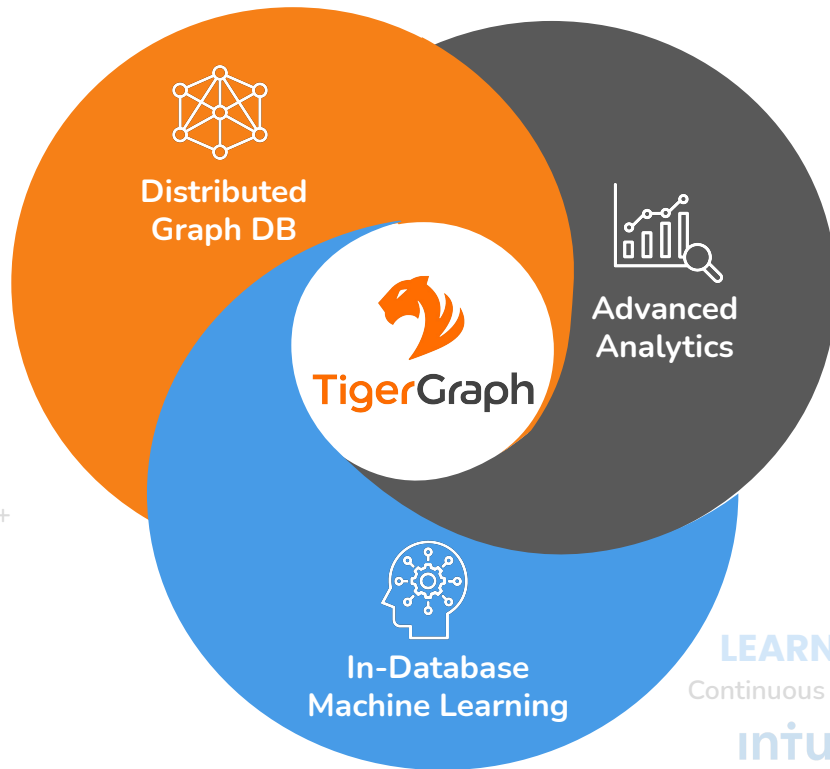
Customer 360 connecting 200+ datasets and pipelines

Fortune 50 Retailer

Item 360 for eCommerce across 100+ datasets



Identity graph connecting multiple data pipelines



ANALYZE CONNECTED DATA

10-100X faster than current solutions

Jaguar Land Rover

Supply chain planning - 3 weeks to 45 minutes



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

Automotive: Analyze Supply Chain & Demand Factors

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)



Harry Powell
Director of Data
and Analytics



Alice Grout-Smith
Data Scientist



Martin Brett
Senior Data Architect



Hazel Scourfield
Data Scientist

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

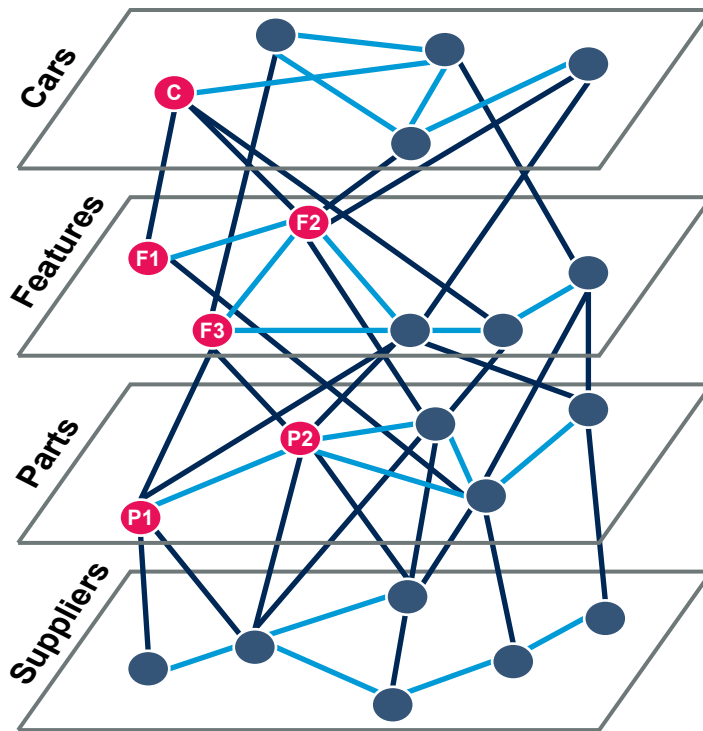
Clear Two-Way Line of Sight Between Demand and Supply

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

Show dependencies
between Suppliers →
Parts → Features → Cars

Include costs and
constraints

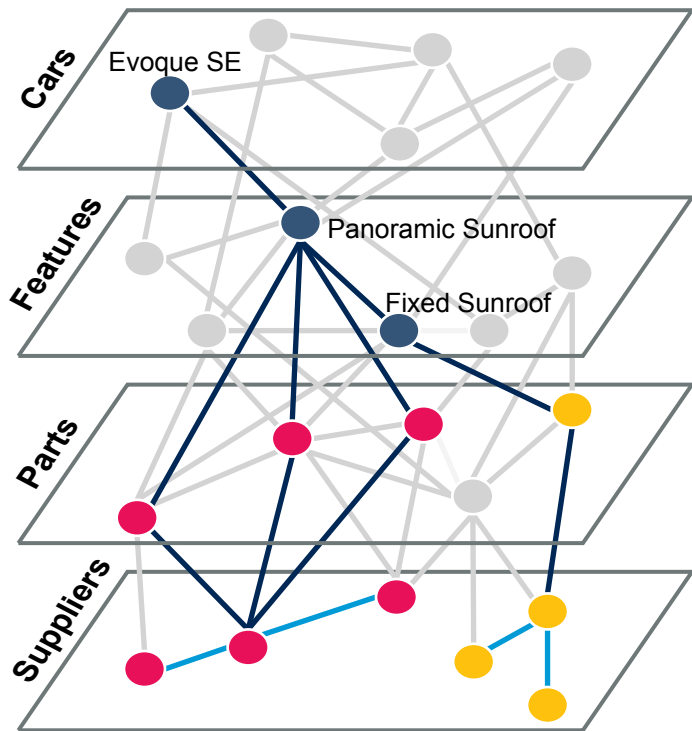
Supply

Source: Adapted From Jaguar Land Rover



Identifying and Reducing Supply Chain Risks

JLR's Demand-Supply Graph for Exploration & Discovery

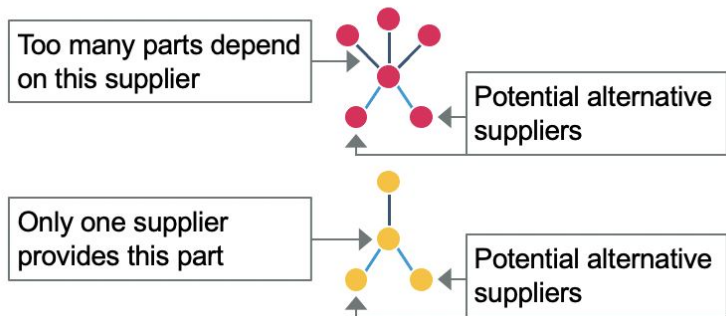


Source: Adapted From Jaguar Land Rover

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?

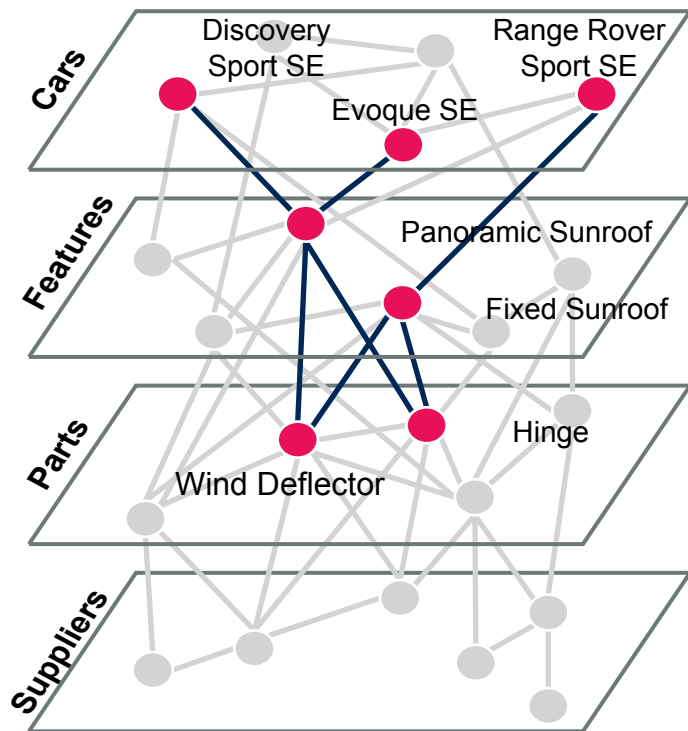
Procedure for Exploration:

- 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.



Making the Most of Surplus Inventory

JLR's Demand-Supply Graph for Investigation & Inference

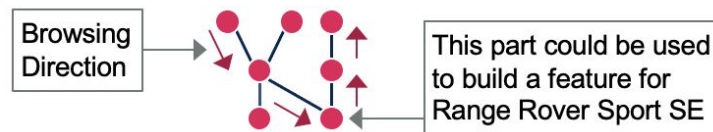


Source: Adapted From Jaguar Land Rover

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?

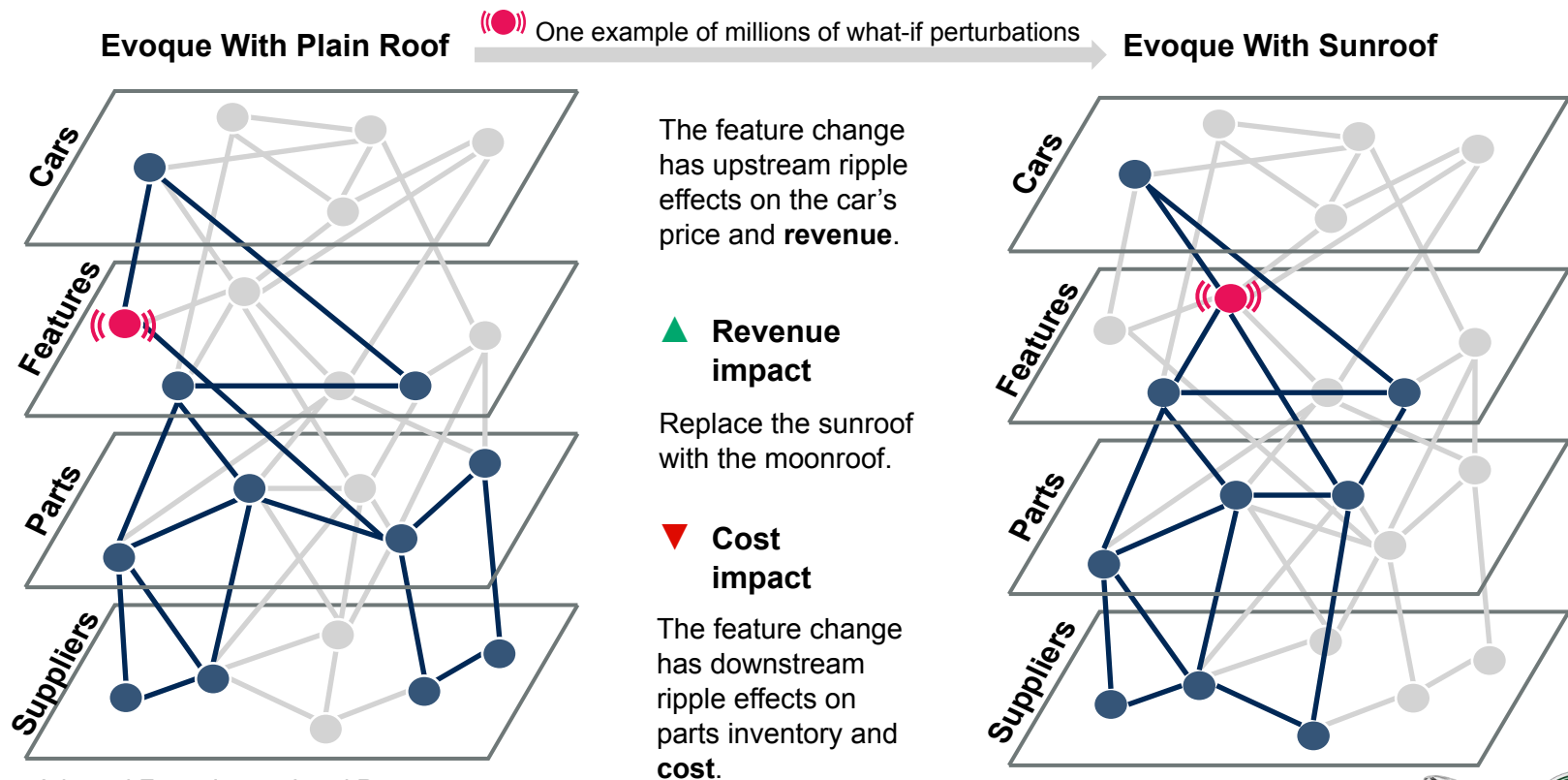
Procedure for Investigation:

- 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.



Solving an Intractable Optimization Problem

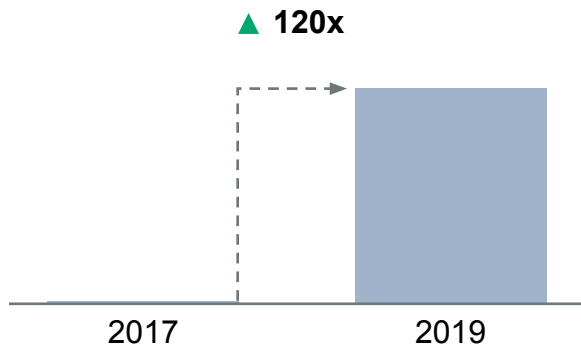
Critical Business Question: What is the profitability impact of changing a feature of a car?



Source: Adapted From Jaguar Land Rover

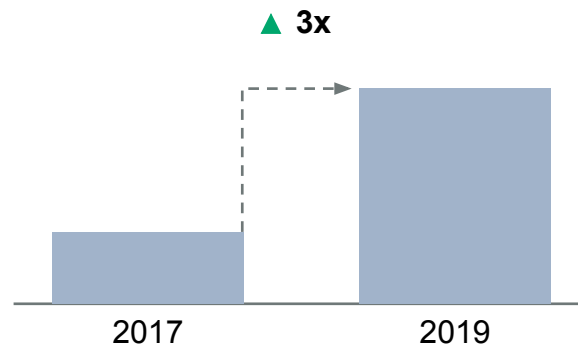
Results

Decision Speed



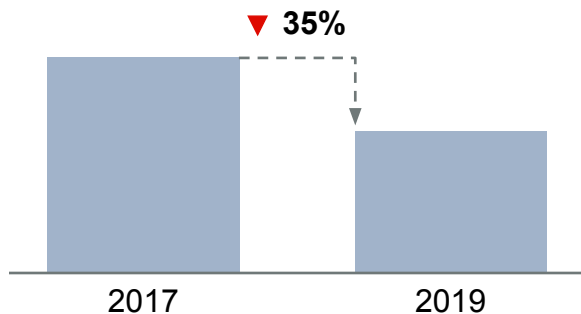
Source: Adapted From Jaguar Land Rover

Business Value



Source: Adapted From Jaguar Land Rover

Supplier Risk



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.

***"The pandemic pivot: IT leaders innovate on the fly",
CIO.com, 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 PB with lowest cost of ownership

UNITEDHEALTH GROUP*

Customer 360 connecting 200+ datasets and pipelines

Fortune 50 Retailer

Item 360 for eCommerce across 100+ datasets



Identity graph connecting multiple data pipelines



Distributed Graph DB



Advanced Analytics



In-Database Machine Learning

ANALYZE CONNECTED DATA

10-100X faster than current solutions

Jaguar Land Rover

Supply chain planning - 3 weeks to 45 minutes



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

Fraud Detection: Need for Better AI

**\$30B
Lost to
Fraud**

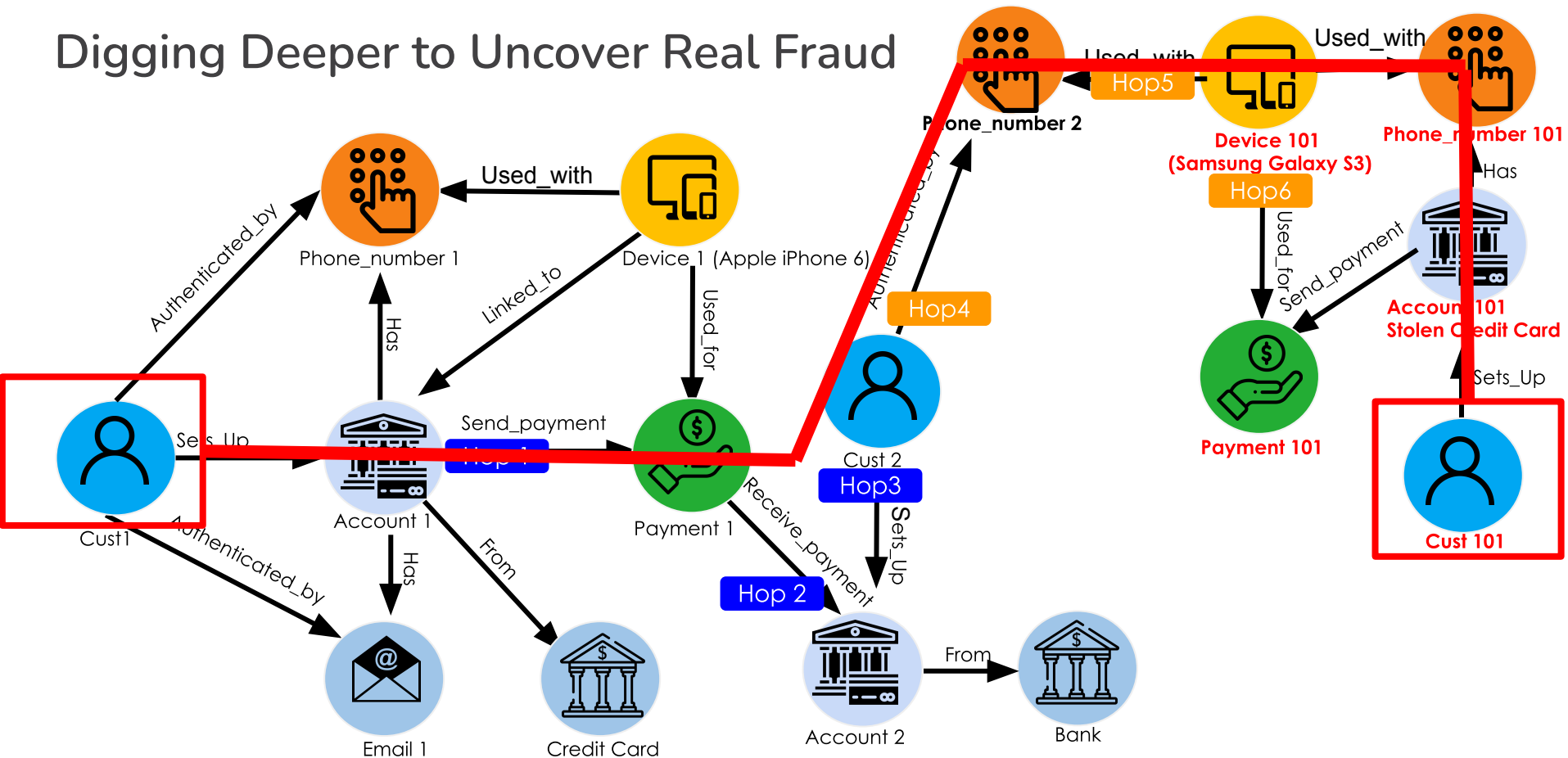


- \$118 billion of blocked sales in the U.S. with 15% of cardholders experiencing blocked sales
- High-income consumers (> \$75,00/yr) at higher risk of false positives (22%)
- 40% of denied users are attempting to pay a greater than \$250 transaction

**80% false
positives
(blocking
non-fraud
transactions)**



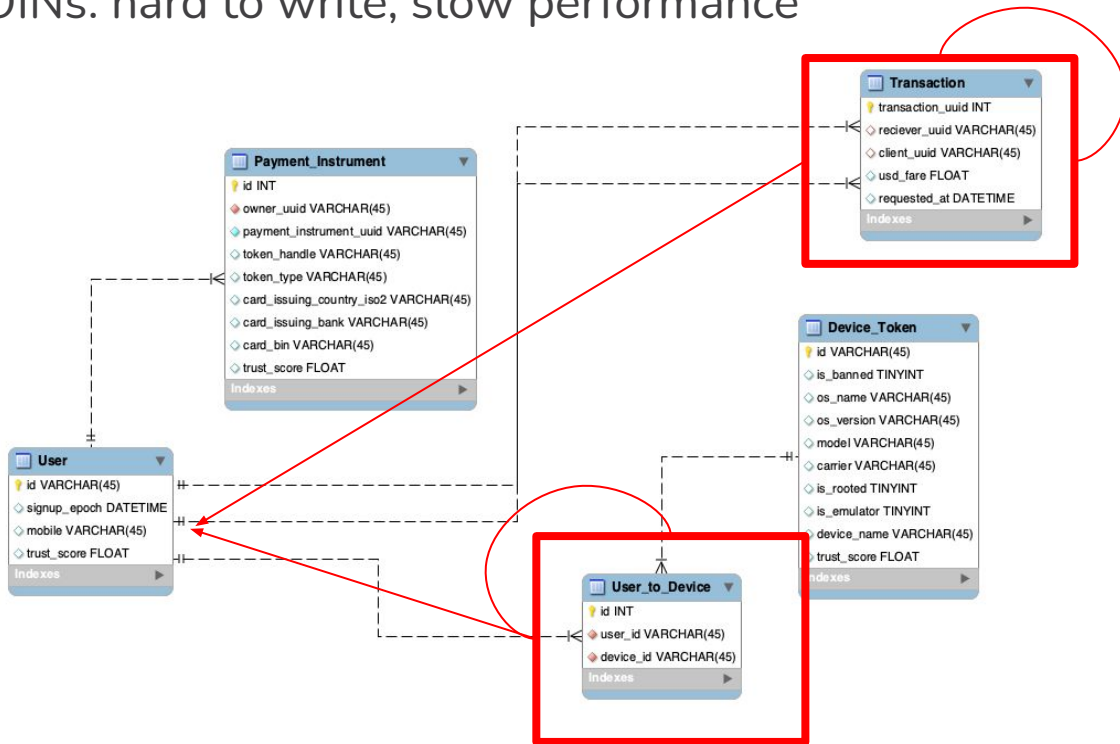
Digging Deeper to Uncover Real Fraud



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

RDBMS Can't Support Real-Time Traversal of Connected Data

Complex JOINS: hard to write, slow performance



Preventing Fraudulent Loans with TigerGraph



Business Challenge

Leading U.S. bank needed to search **20TB** of data for connections between known fraudulent credit card applications and applications of unknown status. Relational databases and other graph providers unable to deliver **speed and scale**.

Solution

- Pair **graph with machine learning** to identify fraud at scale and intervene **in real-time**.
- Leverage **deep link analytics** to find hidden connections across 20TB+ of data.

Business Benefits

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

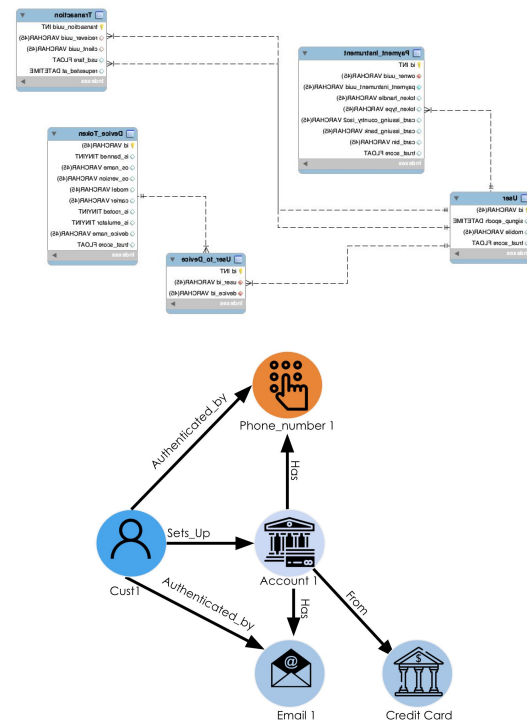
Tier 1 U.S. Bank

- **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

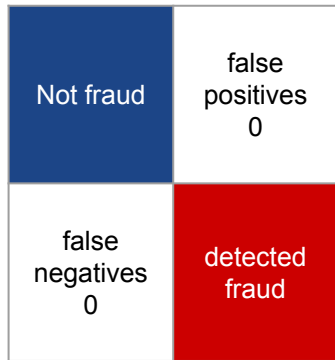
Test Case: Feature Extraction for ML/AI

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

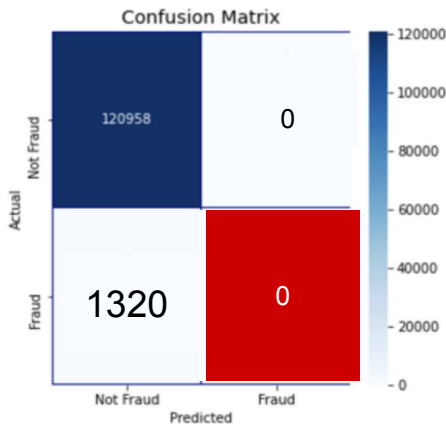


Double the performance of Fraud Detection System with 50% reduction in false positives & half the undetected fraud transactions with Graph DB features when compared to legacy solution

More Accurate Fraud Detection using Graph

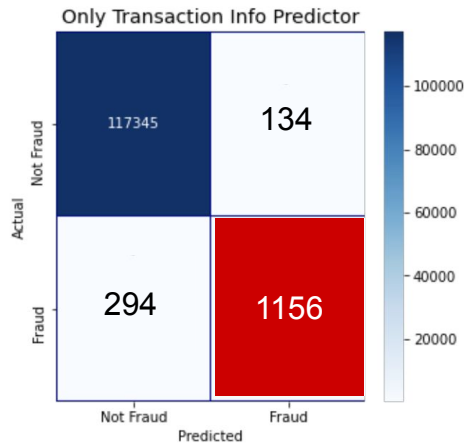


Ideal Fraud Detector



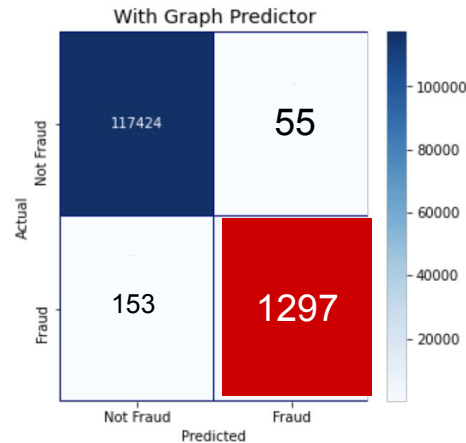
No Fraud Detection:

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

Leading US bank wanted a better way to detect and remove fraudsters from their credit-card network. Prototypes showed that a combination of advanced graph algorithms gave significant gains, but big-data tools and other graph technologies either couldn't scale or gave inconsistent results.

Solution

- Implement **PageRank and Louvain community detection** in an MPP native-parallel database.
- Leverage deep analytics to find hidden connections across 20TB+ of data.

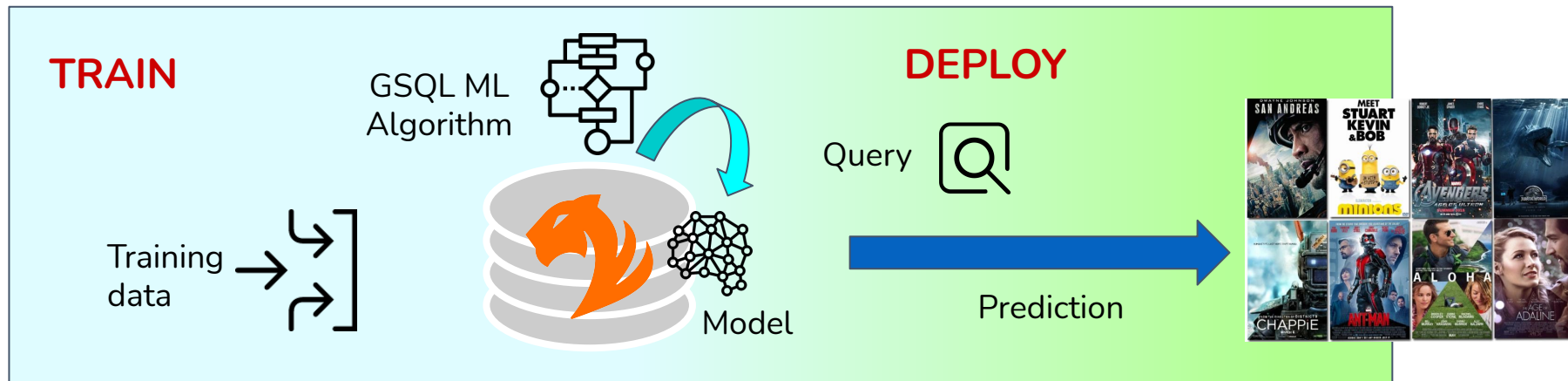
Business Benefits

- 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



- 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
- ...



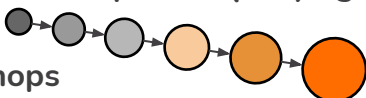



In-Graph Machine Learning

Efficient Machine Learning requires:

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



The TigerGraph Difference

Feature	Design Difference	Benefit
Real-Time Deep-Link Querying  5 to 10+ hops	<ul style="list-style-type: none">• Native Graph design• C++ engine for high performance• Storage Architecture	<ul style="list-style-type: none">• Uncovers hard-to-find patterns• Operational, real-time• HTAP: Transactions+Analytics
Handling Massive Scale 	<ul style="list-style-type: none">• Distributed DB architecture• Massively parallel processing• Compressed storage reduces footprint and messaging	<ul style="list-style-type: none">• Integrates all your data• Automatic partitioning• Elastic scaling of resource usage
In-Database Analytics & Machine Learning 	<ul style="list-style-type: none">• GSQL: High-level yet Turing-complete language• User-extensible graph algorithm library, runs in-DB• ACID (OLTP) & Accumulators (OLAP)	<ul style="list-style-type: none">• Avoids transferring data• Richer graph context• Graph-based feature extraction for supervised machine learning• In-DB machine learning training
	<ul style="list-style-type: none">• No-code migration from RDBMS• No-code Visual Query Builder	<ul style="list-style-type: none">• Democratize self-service analytics to derive new-insights from legacy/external data stores



Starter Kits and Developer Portal for Graph+ML



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



TigerGraph

© 2020. ALL RIGHTS RESERVED. | TIGERGRAPH.COM |

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