

# Real-Time Deep-Link Analytics for Big Graphs

---

## Challenges and Solutions

Victor Lee, Sr. Product Manager



BigGraph Meetup  
October 11, 2017

Welcome to

**BigGraph**

*Meetup*

We meet to network, share, discuss, and invent together graph technologies that empower the analysis capabilities needed for today's most critical enterprise applications.

Goal: real-time and mutable Big Graph data management platform supporting deep link analytics

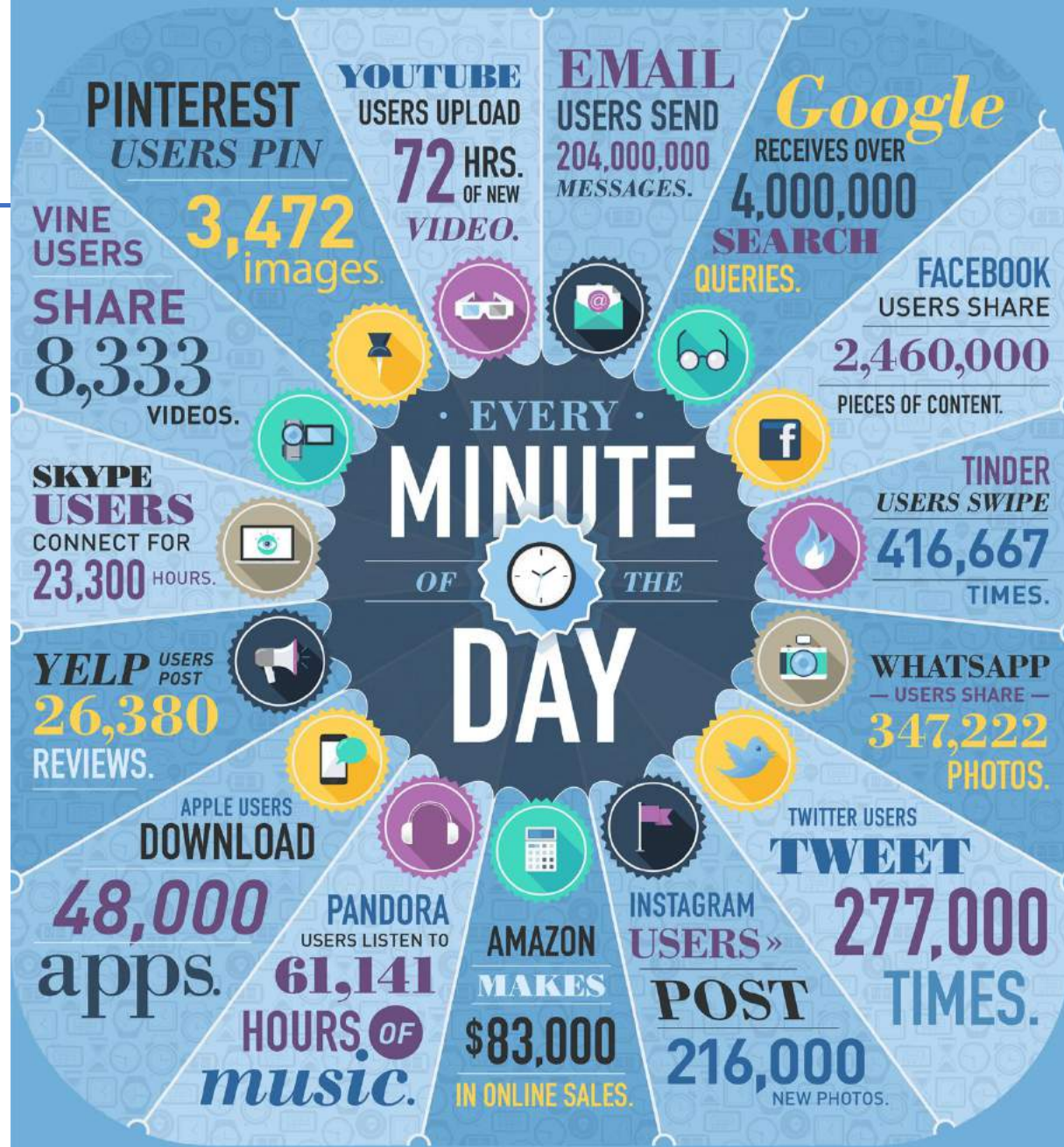
- Big Graph? Billion-scale Graphs, with 10B to 100B+ vertices and/or edges.
- Real-time milli-second query response time on Big Graphs.
- Mutable data mgmt supporting real-life scenarios, > 100K+ updates per second.
- Deep link analytics queries which traverse 3 to 10+ hops deep into the graph.

# The Data Deluge

## Data Never Sleeps 2.0

from Domo.com

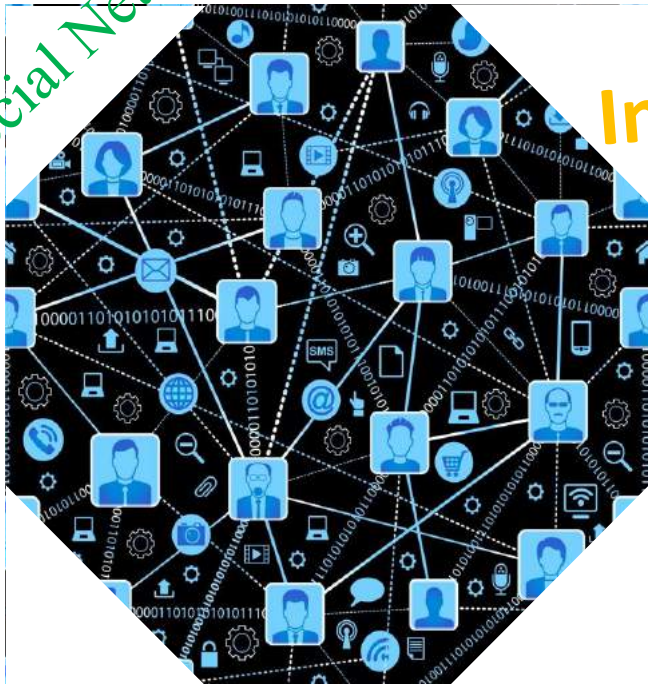
<https://www.domo.com/learn/data-never-sleeps-2>



# The Networked Age



Social Networks



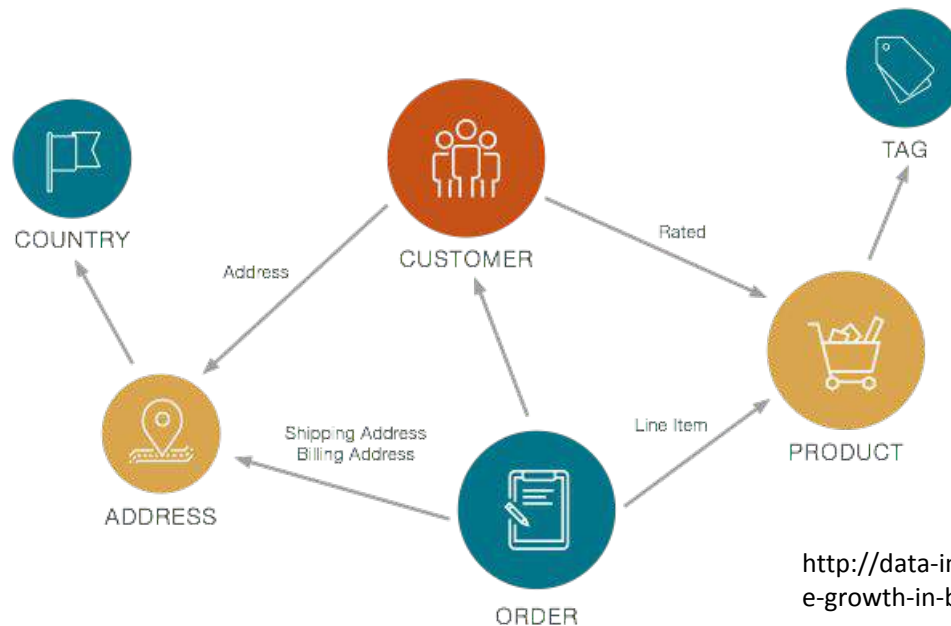
Internet of Things

BIG DATA



# Graph Data Analytics to the Rescue

- Natural storage model for connectional data
- Natural model for many types of transactions
- Natural computational model for knowledge/inference/learning – chaining and combining observations



<http://data-informed.com/graph-databases-se-e-growth-in-business-applications-adoption/>

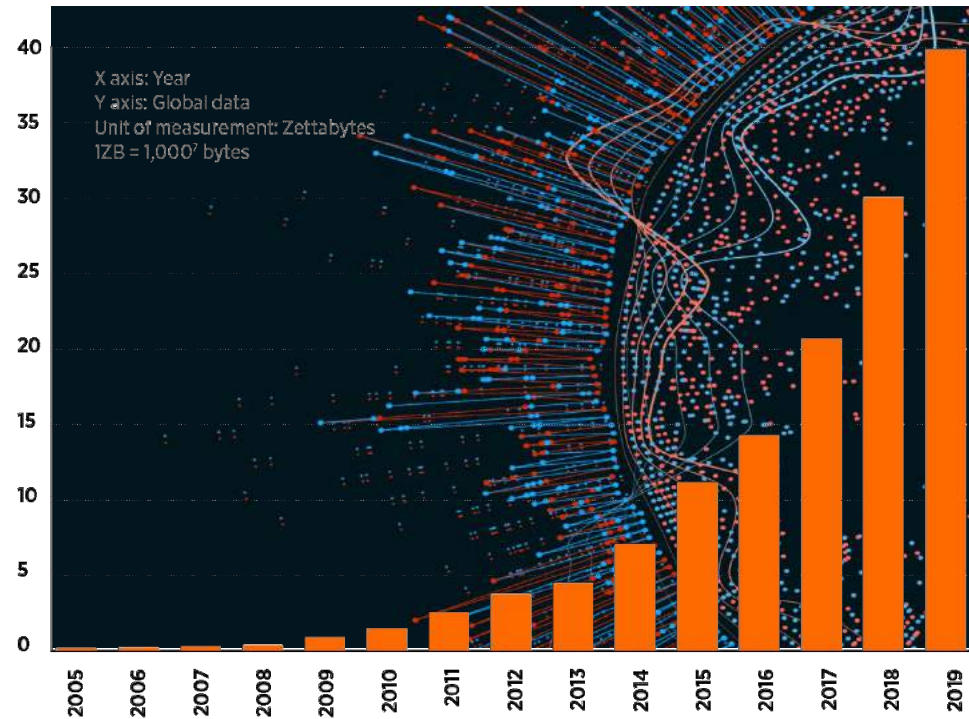
# Challenge: Scalability

- Datasets are getting bigger
- Outgrowing a single server in many cases.
- More data → better models/predictions/results
- **Problem:** Some graph DBs cannot distribute across multiple nodes.

## Requirement:

Data Platform must scale-up and scale-out well

## DATA GROWTH



Note: Post-2013 figures are predicted. Source: UNECE

<https://www.theneweconomy.com/strategy/big-data-is-not-without-its-problems>

# Challenge: Real-Time Processing

Many applications require real-time processing.  
Gets harder as data sizes get bigger.



- Processing transactions
- Engaging customers
- Catching fraud in progress
- Managing live systems
- Ingesting streaming data

## **Requirement:**

Speed needed both for  
querying and data  
updates/loading

# Challenge: Deep Link Analytics

- Each additional hop in a graph reveals more information / knowledge / evidence
- Deeper traversal → better models/predictions/results
- **Problem:** Most graph DBs slow down or timeout after 2 hops.

## Requirement:

Data Platform must traverse multiple hops efficiently





# Need: Scalable Real-Time Deep Link Analytics

## Deep Link Analytics: Queries With 3 Or More Hops



Deep Link Analytics are essential for enterprise applications including AI and machine learning apps

fraud and risk management, anti-money laundering  
personalized recommendation, knowledge graph, customer identity graph, supply-chain logistics, ...



Requires a real-time graph platform that shatters current graph database speed and performance limitations

# Solution: Native Parallel Graph

## TigerGraph: The First Native Parallel Graph

Designed From the Ground Up For Computational Parallelism



**NATIVE**  
graph storage



**PARALLEL GRAPH**  
computation engine



**10X+**  
compression

# Native Parallel Graph

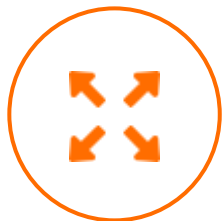
## Automatic Computational Parallelism



Each vertex/edge is not only a unit of storage, but is also a computational unit

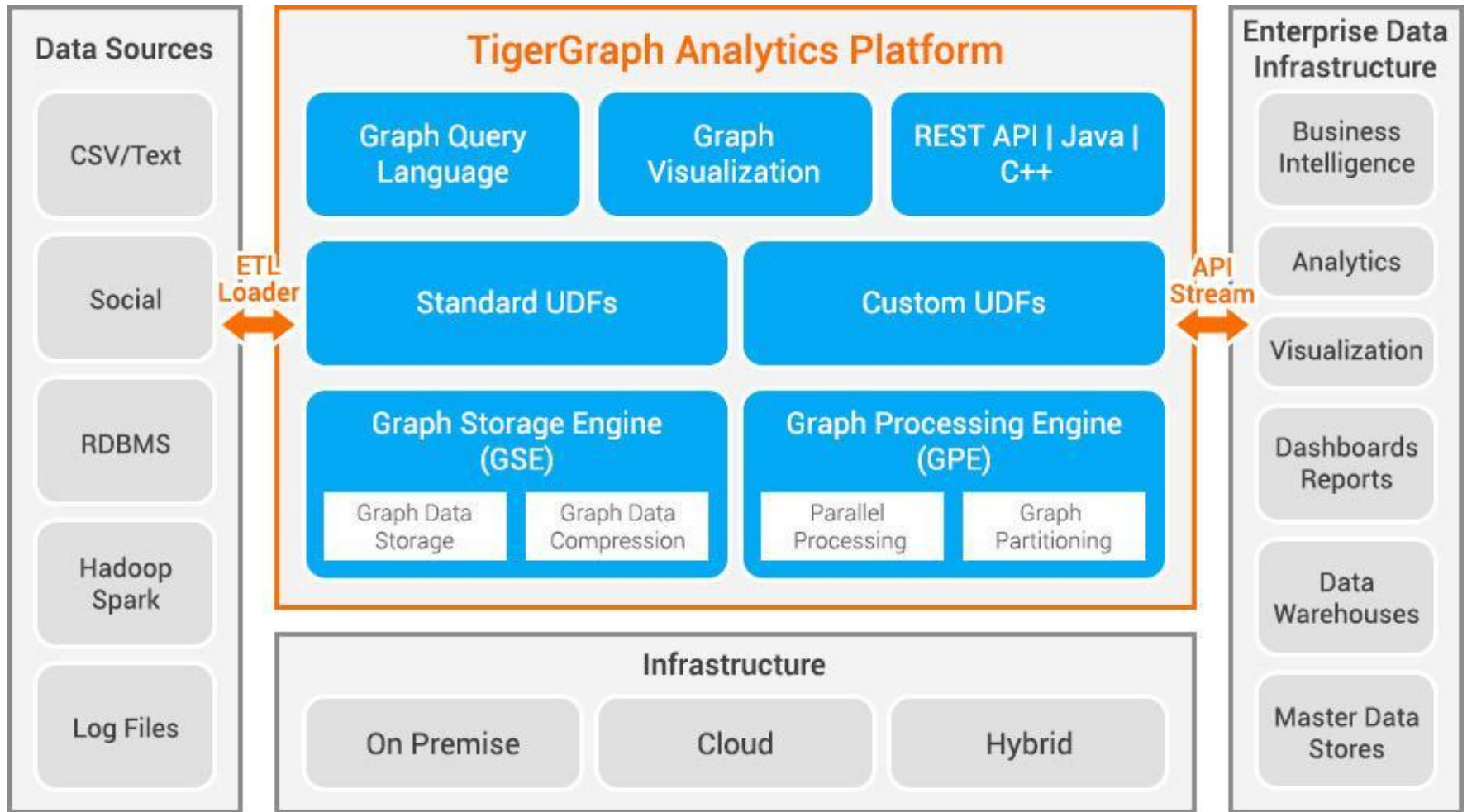


Each vertex/edge is processed in parallel



TigerGraph engine will automatically scale the computation across all threads and CPU cores available

# Real-Time Graph Analytics Platform



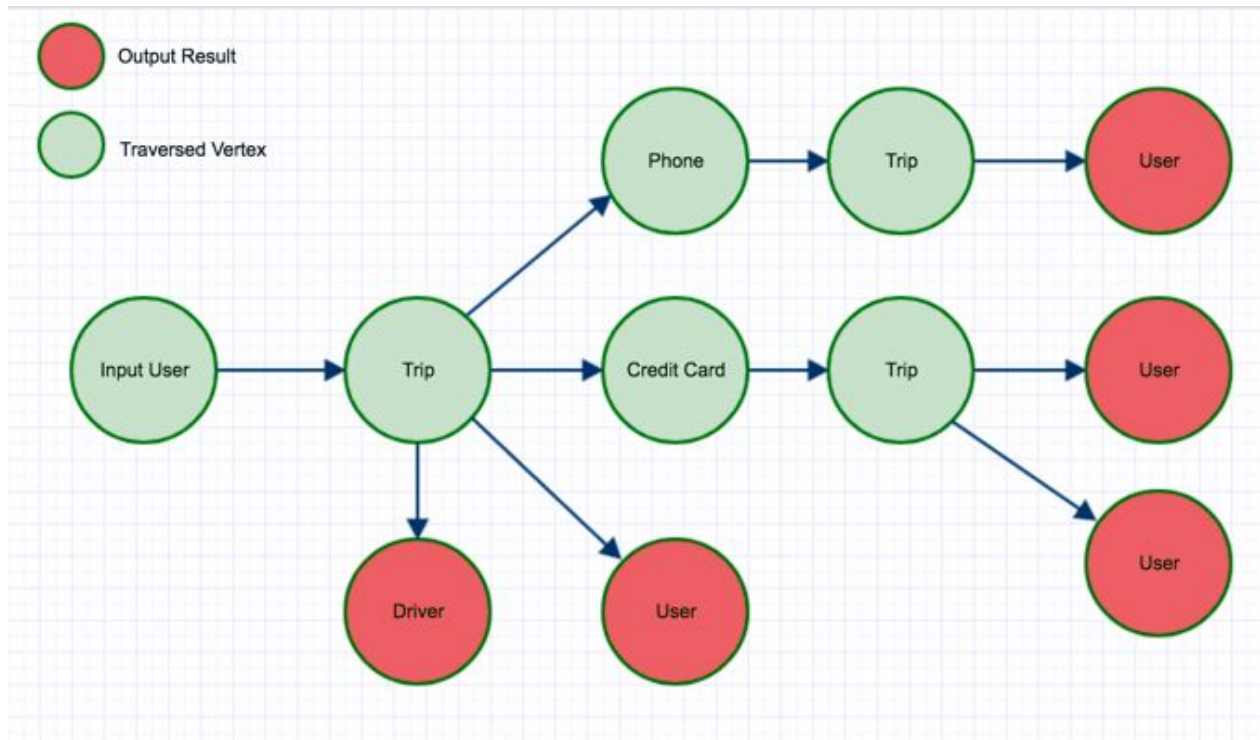
# Demo: Queries Per Second on a Big Graph

---

- Graph Schema and Size: Delivery service
  - 7 vertex types: Order, User, Location, CreditCard, etc.
  - 9 edge types: Order\_User, Order\_Location, Order\_Card, etc.
  - 3.4 Billion vertices
  - 22.5 Billion edges
- Test Server
  - Amazon EC2 i3.16xlarge
  - 64 vCPU, 488 GiB Memory, 15 TB SSD Storage

# Demo: Queries Per Second on a Big Graph

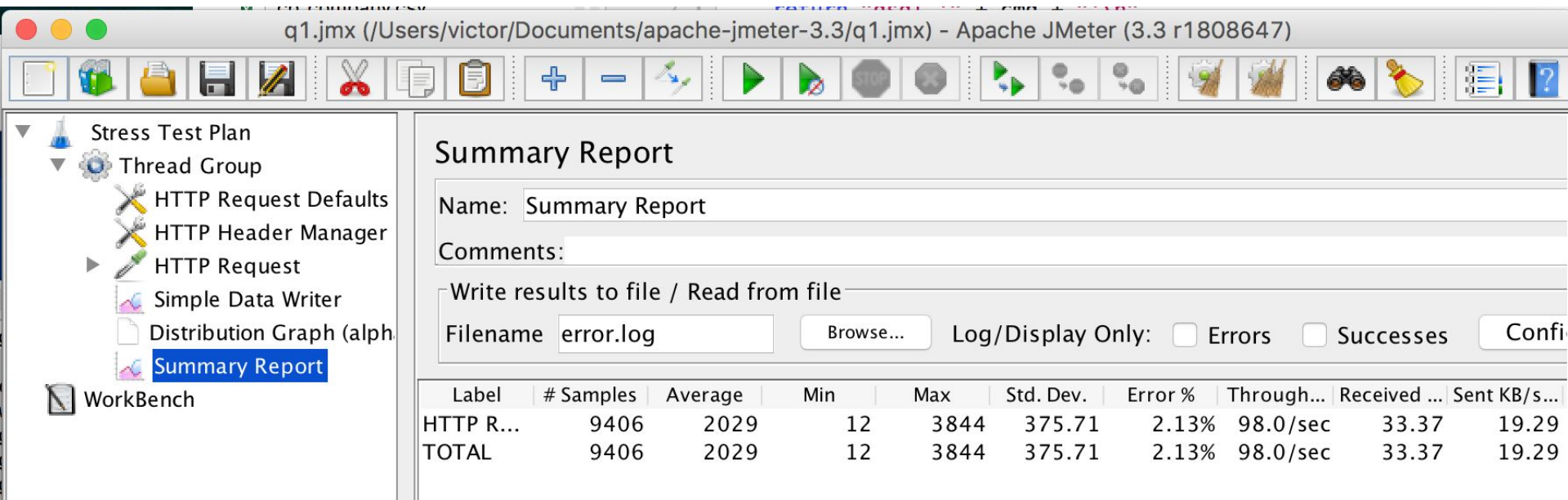
Starting with a Driver\_id X, find all other Driver\_ids/User\_ids connected to the Driver\_id X through phone\_number, credit\_card connected via trips in 4 steps



# Demo: Queries Per Second on a Big Graph

Based on the graph density, each query can traverse about 100K edges.  
The live test shows that we can process 98.0 queries per second per machine  
→ 10M edges per second, for a 4-hop query on a graph with 22B edges.

Even better speed is possible for a less demanding test.



The screenshot shows the Apache JMeter interface with a Summary Report for a test plan. The report displays the following data:

Label	# Samples	Average	Min	Max	Std. Dev.	Error %	Through...	Received ...	Sent KB/s...
HTTP R...	9406	2029	12	3844	375.71	2.13%	98.0/sec	33.37	19.29
TOTAL	9406	2029	12	3844	375.71	2.13%	98.0/sec	33.37	19.29

# A Closer Look At the TigerGraph System

---



web browser based tool to visually and interactively define,  
create, explore and query big graphs

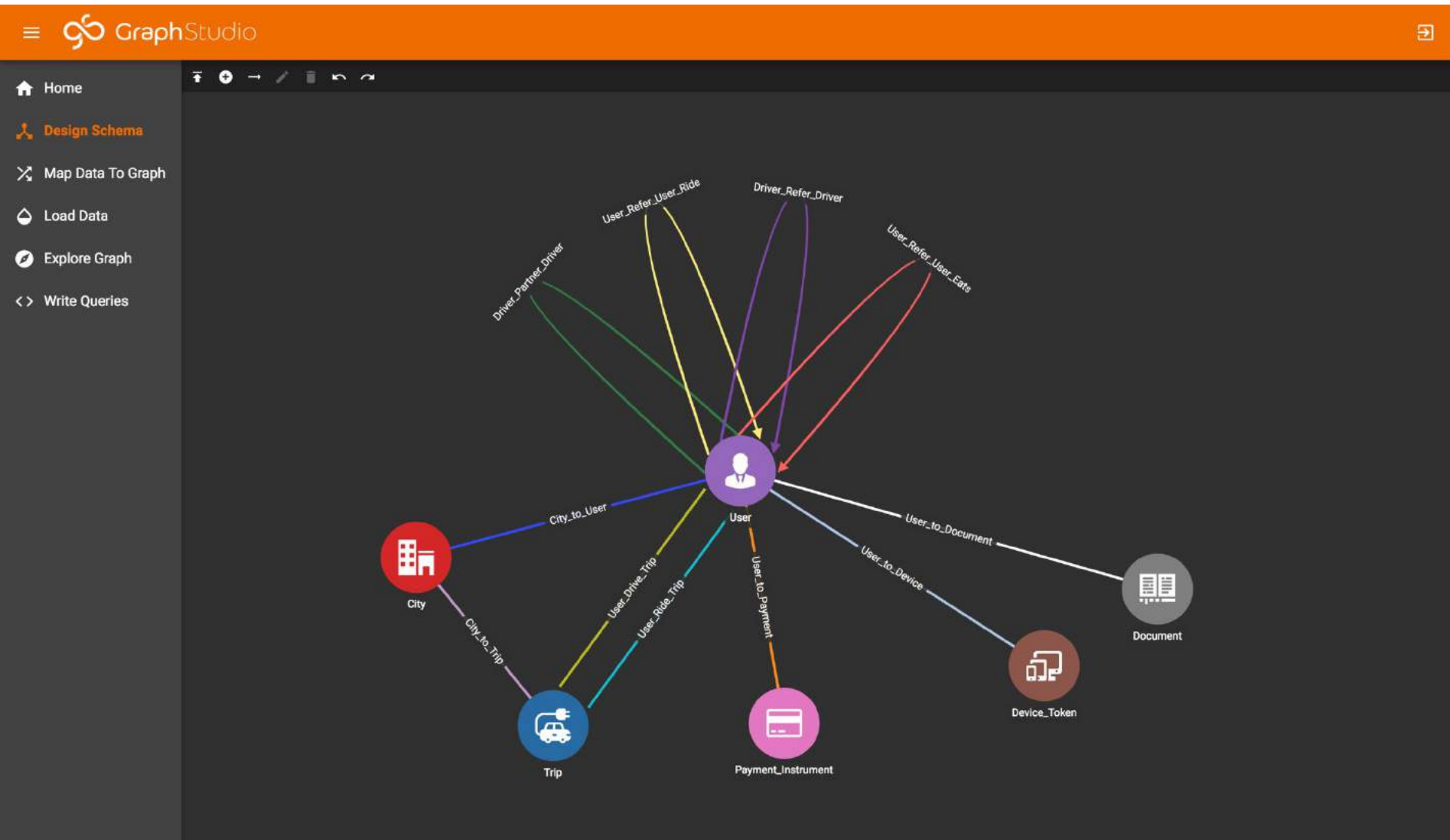


# GraphStudio: Visual Graph Schema Designer

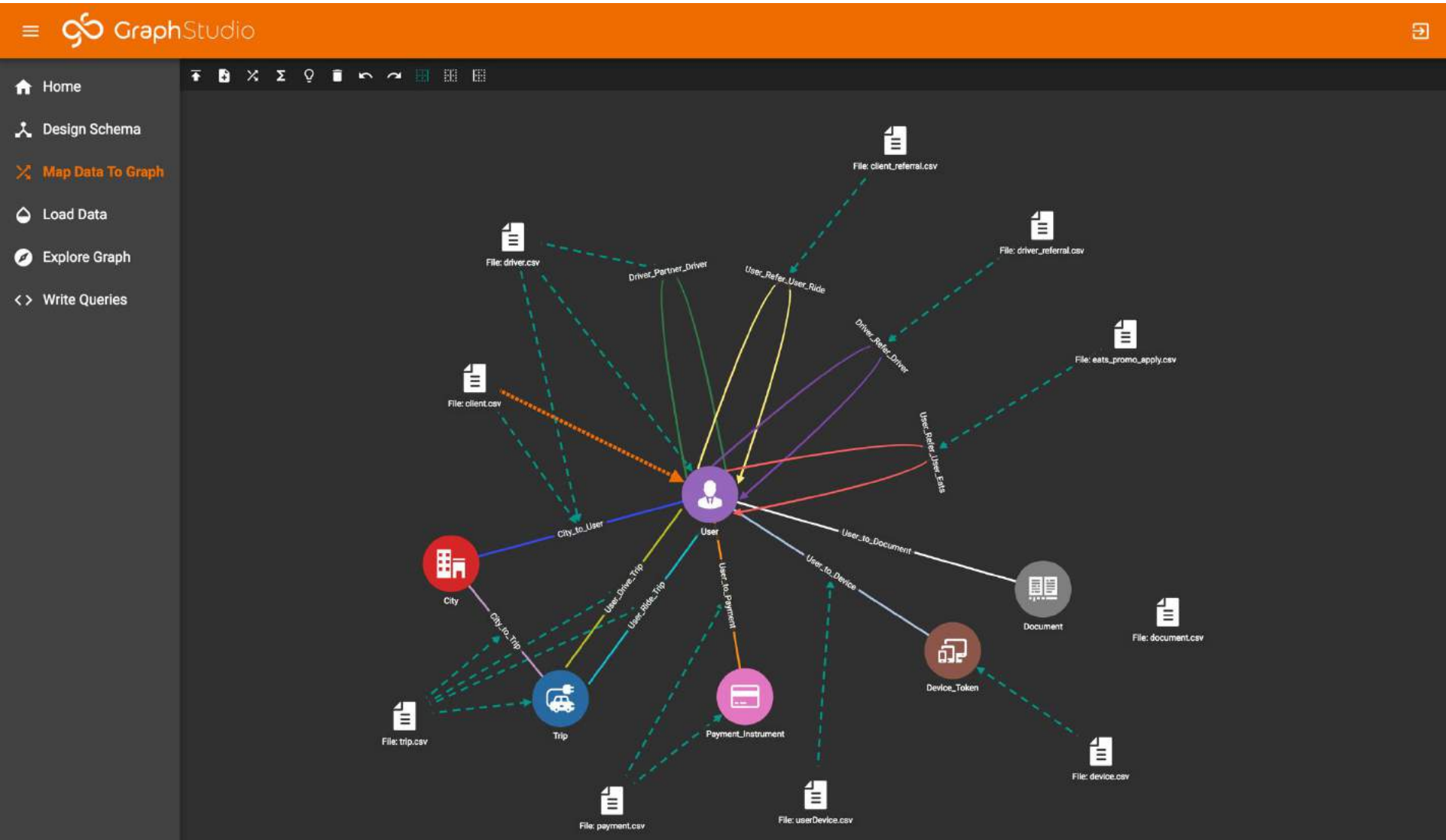
The screenshot displays the GraphStudio interface. On the left is a sidebar with navigation options: Home, Design Schema (highlighted), Map Data To Graph, Load Data, Explore Graph, and Write Queries. The main workspace shows a graph schema with three vertices: Person (blue circle with a person icon), Enterprise (blue circle with a building icon), and Project (orange circle with a group of people icon). Edges connect these vertices: 'person\_invest\_enterprise' (dashed line), 'work\_for' (dashed line), 'bid\_for' (solid orange line), and 'enterprise\_invest\_enterprise' (dashed line). A tooltip above the Person vertex reads 'Select the target vertex to create an edge'. A table below the Project vertex shows its schema:

Vertex	Project
(PRIMARY ID) name	STRING

# GraphStudio: Visual Graph Schema Designer



# GraphStudio: Visual Graph Data Loader





# GraphStudio: Graph Explorer

The screenshot displays the GraphStudio Graph Explorer interface. The top navigation bar includes a hamburger menu, the GraphStudio logo, and a search icon. The left sidebar contains navigation options: Home, Design Schema, Map Data To Graph, Load Data, Explore Graph (highlighted), and Write Queries. The main panel shows a graph visualization with nodes representing people (e.g., Joseph, Sarah, Robert, Susan, James, Richard, Carol, Mary, Paul) and projects (e.g., Vibber Torch, HuskyOople, Umbrella Telecom Fraud Detection, PepperSrik, PepperPipe, ArgonGraph, Wipe Movie Recommendation). Edges represent relationships like 'work\_for', 'person\_invest\_enterprise', and 'bid\_for'. The sidebar also features a search section for vertices and a configuration panel for search filters.

**Search Vertices In Graph**

Enter vertex id  Search

Enter a number  **Pick Vertices**

**Configuration**  
Search within vertex types

- All
- Enterprise
- Person
- Project

# GraphStudio: Graph Explorer

The screenshot displays the GraphStudio Graph Explorer interface. On the left is a navigation sidebar with options: Home, Design Schema, Map Data To Graph, Load Data, Explore Graph (highlighted), and Write Queries. The main area is divided into a configuration panel on the left and a graph visualization on the right.

**Configuration Panel:**

- Find Connections Between Multiple Vertices:** Click vertices in right panel to choose vertices. Press "Shift" key to choose multiple vertices at same time.
- Find Connection Paths:** Button to initiate path finding.
- Configuration:**
  - Maximal connection path length: 10
  - Warning: Finding connection paths longer than 6 steps might be slow.
- Connected by vertex types:**
  - All
  - Trip
  - City
  - User
  - Device\_Token
  - Payment\_Instrument
  - Document
- Connected by edge types:**
  - All
  - User\_Drive\_Trip
  - User\_Ride\_Trip
  - User\_to\_Device
  - User\_to\_Payment
  - City\_to\_Trip
  - User\_to\_Document
  - User\_Refer\_User\_Ride
  - Driver\_Refer\_Driver
  - User\_Refer\_User\_Eats
  - City\_to\_User
  - Driver\_Partner\_Driver

**Graph Visualization:** A complex graph with nodes representing users, cities, and trips. Nodes are color-coded: purple for users, blue for cities, and red for trips. Edges represent relationships between these entities. A path of length 10 is highlighted in yellow, starting from a red trip node and ending at a purple user node.

# GraphStudio: Graph Query Editor

The screenshot displays the GraphStudio interface, which is divided into several sections:

- Header:** Features the GraphStudio logo and a search icon.
- Left Sidebar:** Contains navigation options: Home, Design Schema, Map Data To Graph, Load Data, Explore Graph, and Write Queries (highlighted in orange).
- GSQL Queries Panel:** Lists several queries: GrossBooking, MultiTrip (selected), RepeatedUser, and SameDriverRider.
- Enter Query Parameters Dialog:** A modal window for the selected query with input fields for:
  - endDate: int number (value: 2000000000)
  - step: int number (value: 10)
  - startDate: int number (value: 0)A "Run Query" button is located at the bottom.
- Query Editor:** Displays the following GSQL code:

```
1 - CREATE QUERY MultiTrip (VERTEX<Trip> trip, int step, int startDate=0, int endDate=1999999999) FOR GRAPH
2   poc_graph {
3     SetAccum<VERTEX> @@tripSet;
4     OrAccum<bool> @fromDriver, @fromRider;
5     SetAccum<VERTEX> @@userSet;
6     SetAccum<EDGE> @@edgeSet;
7
8     Start (ANY) = {trip};
9
10
11
12 Start = SELECT t FROM Start:s-((User_Drive_Trip|User_Ride_Trip):e)-:t
13   ACCUM
14     case when e.type == "User_Drive_Trip" then
15       t.@fromDriver += true
16     else
17       t.@fromRider += true
18     end
19     ,@@userSet += t
20     ,@@edgeSet += e
21 ;
```
- Graph Visualization:** A large network graph with nodes of various colors (blue, pink, orange) and edges connecting them. A toolbar with various icons is visible above the graph.

# TigerGraph is Hiring!



- Senior Technical Sales Engineer
- Senior Technical Product Manager

\$3000 Referral Bonus! Contact us [talent@tigergraph.com](mailto:talent@tigergraph.com)

See all our open positions at [www.tigergraph.com/join-us/](http://www.tigergraph.com/join-us/)



# Thank you!

---



# TigerGraph

Follow us on Twitter or LinkedIn



to keep up with new developments and opportunities

More questions?

[victor@tigergraph.com](mailto:victor@tigergraph.com)

[info@tigergraph.com](mailto:info@tigergraph.com)