Real-Time Deep-Link Analytics for Big Graphs

Challenges and Solutions

Victor Lee, Sr. Product Manager

TigerGraph

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.

© 2017 TigerGraph. All Rights Reserved.

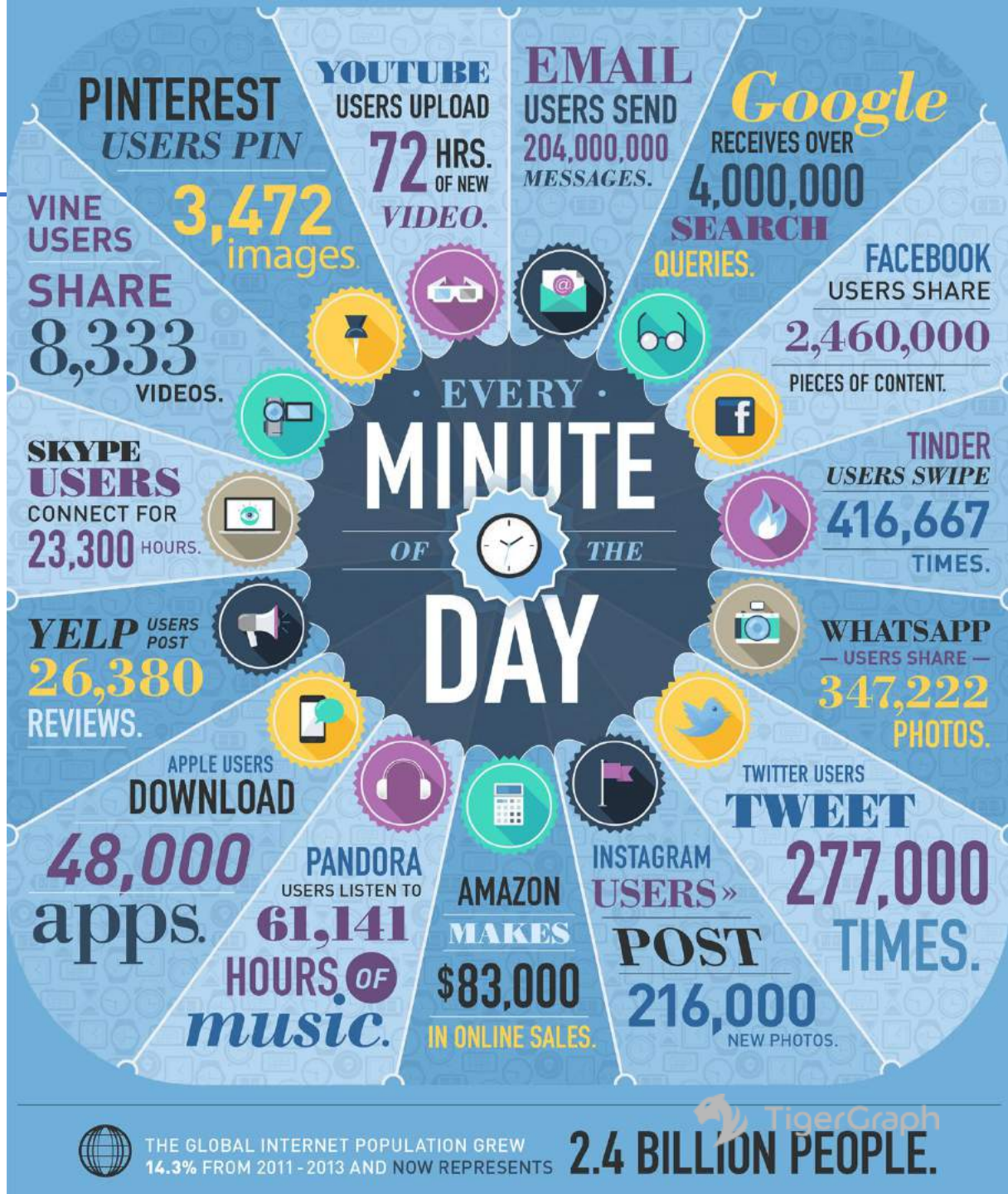
TigerGraph

The Data Deluge

Data Never Sleeps 2.0

from Domo.com

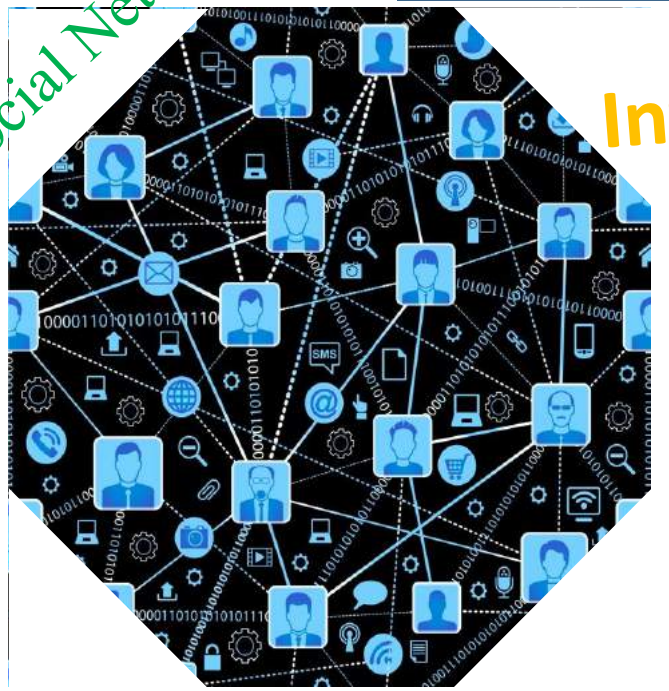
https://www.domo.com/learn/data-never-sleeps-2



© 2017 TigerGraph. All Rights Reserved.

The Networked Age

Social Networks



Internet of Things

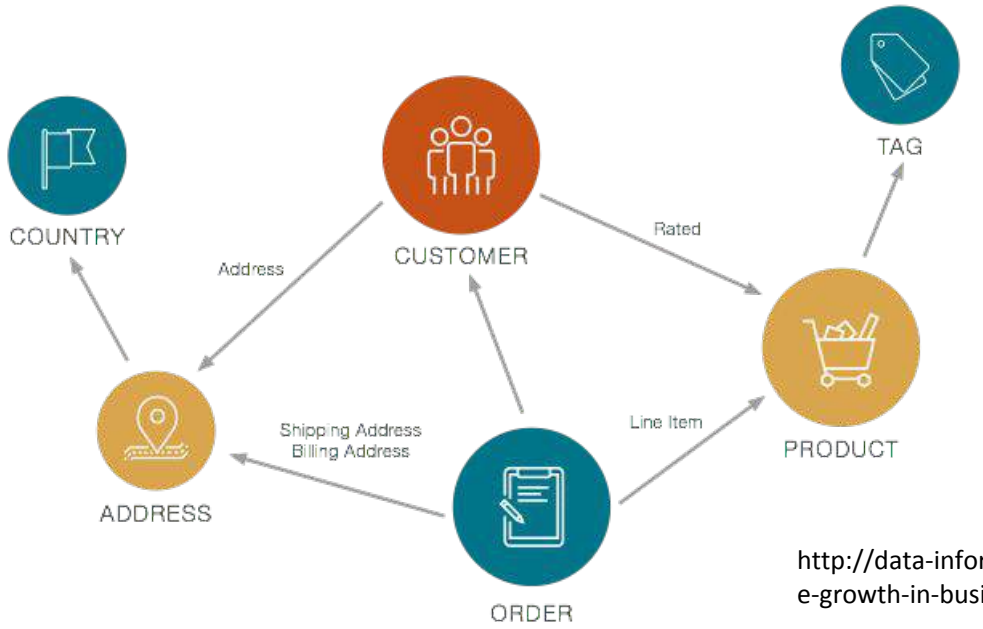
BIG DATA

© 2017 TigerGraph. All Rights Reserved.

TigerGraph

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-see-growth-in-business-applications-adoption/

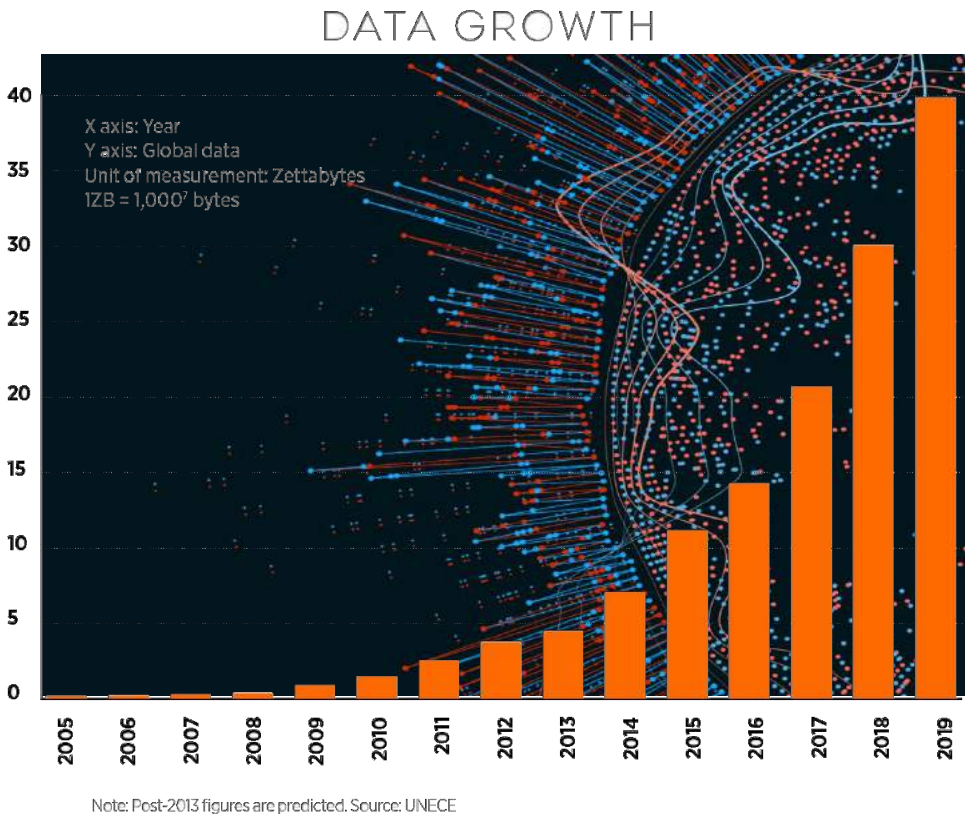
© 2017 TigerGraph. All Rights Reserved.

TigerGraph

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



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

© 2017 TigerGraph. All Rights Reserved.

TigerGraph

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

© 2017 TigerGraph. All Rights Reserved.

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

© 2017 TigerGraph. All Rights Reserved.

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

© 2017 TigerGraph. All Rights Reserved.

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

© 2017 TigerGraph. All Rights Reserved.

Native Parallel Graph

Automatic Computational Parallelism

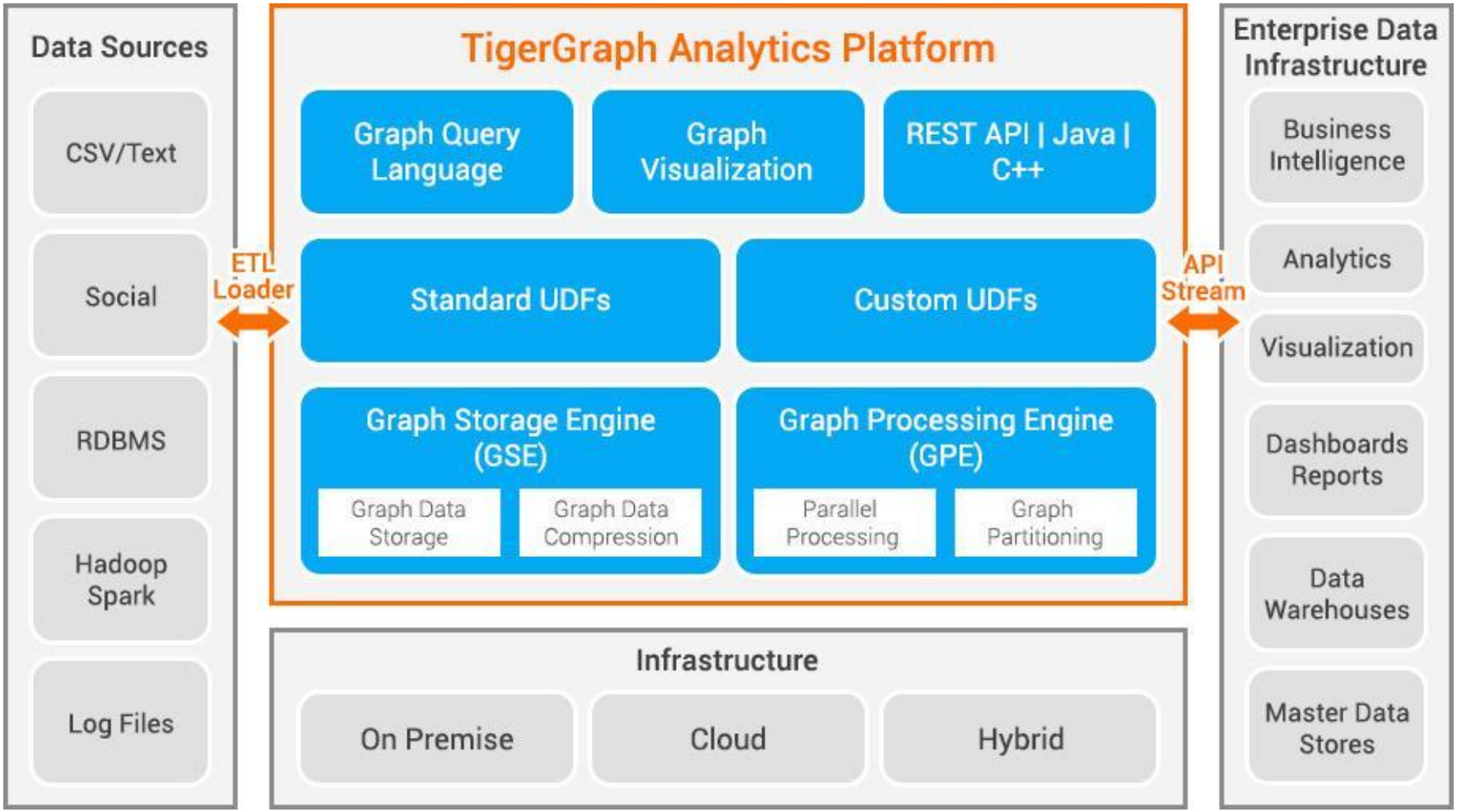
Each vertex/edge is not only a unit of storage, but is also a computational unit

Each vertex/edge is processed in parallel

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

© 2017 TigerGraph. All Rights Reserved.

Real-Time Graph Analytics Platform



© 2017 TigerGraph. All Rights Reserved.

TigerGraph

Demo: Queries Per Second on a Big Graph

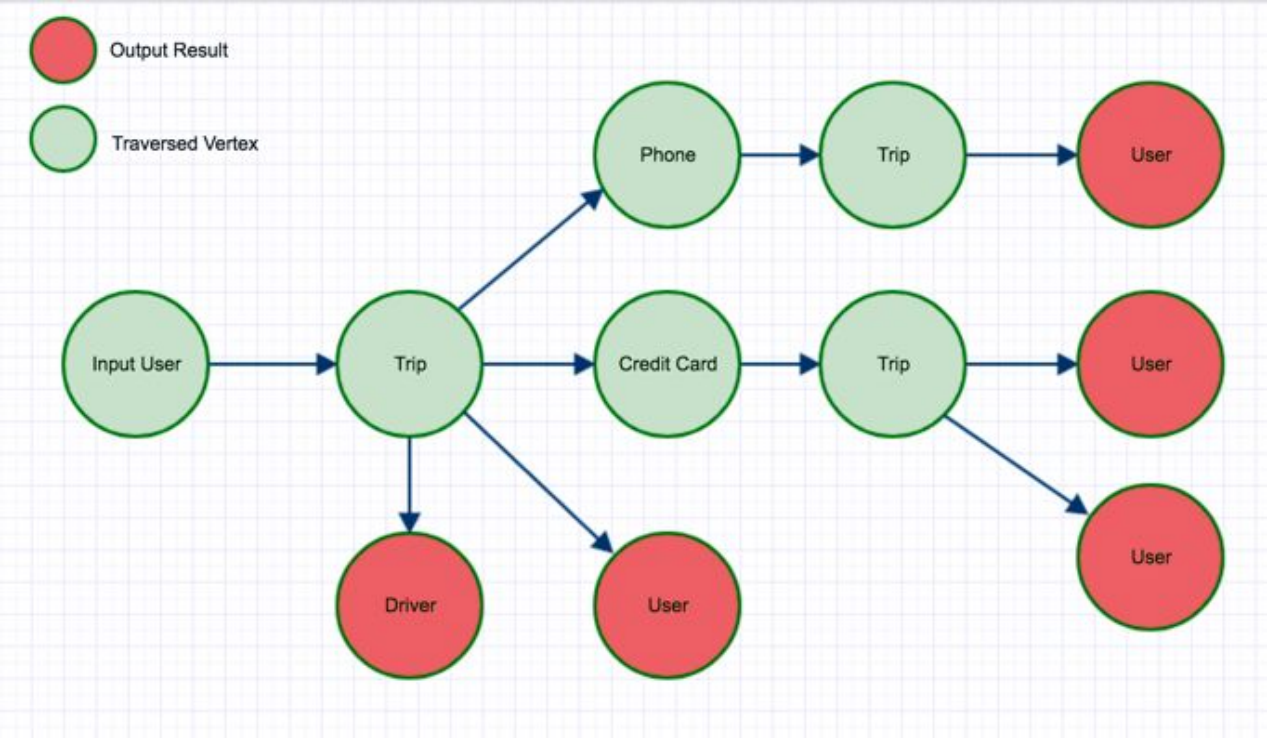
- Graph Schema and Size: Delivery service
  - 7 vertex types: Order, User, Location, CreditCard, etc.
  - 9 edge types: Order_User, Order_Location, Order_Card, etc.
  - 3.4 Billion vertices
  - 22.5 Billion edges

- Test Server
  - Amazon EC2 i3.16xlarge
  - 64 vCPU, 488 GiB Memory, 15 TB SSD Storage

© 2017 TigerGraph. All Rights Reserved.

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



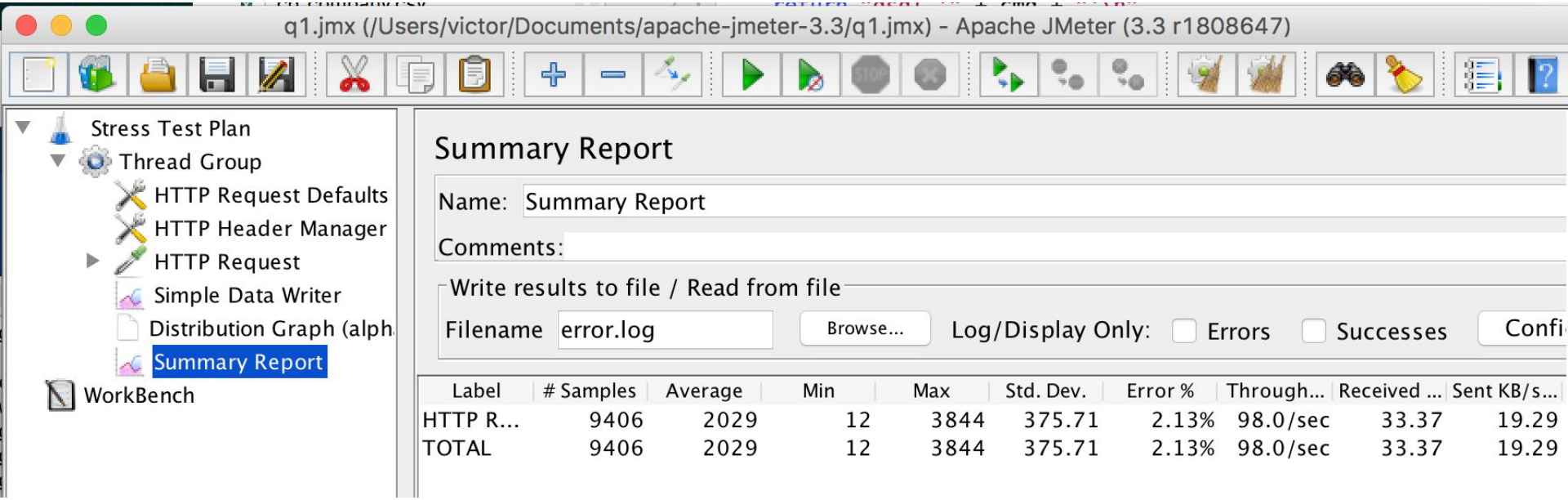
© 2017 TigerGraph. All Rights Reserved.

TigerGraph

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.



© 2017 TigerGraph. All Rights Reserved.

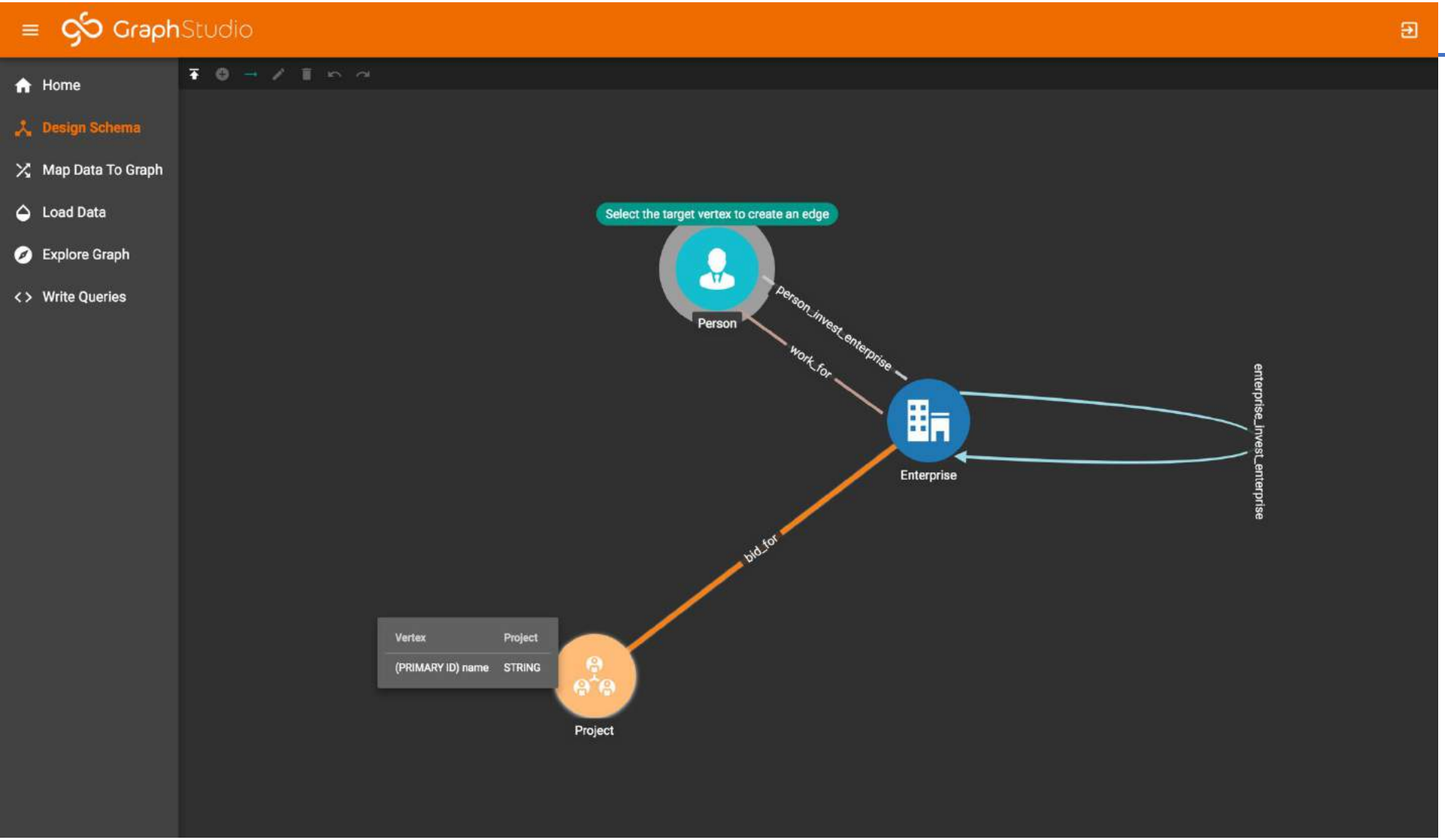
TigerGraph

A Closer Look At the TigerGraph System

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

© 2017 TigerGraph. All Rights Reserved.

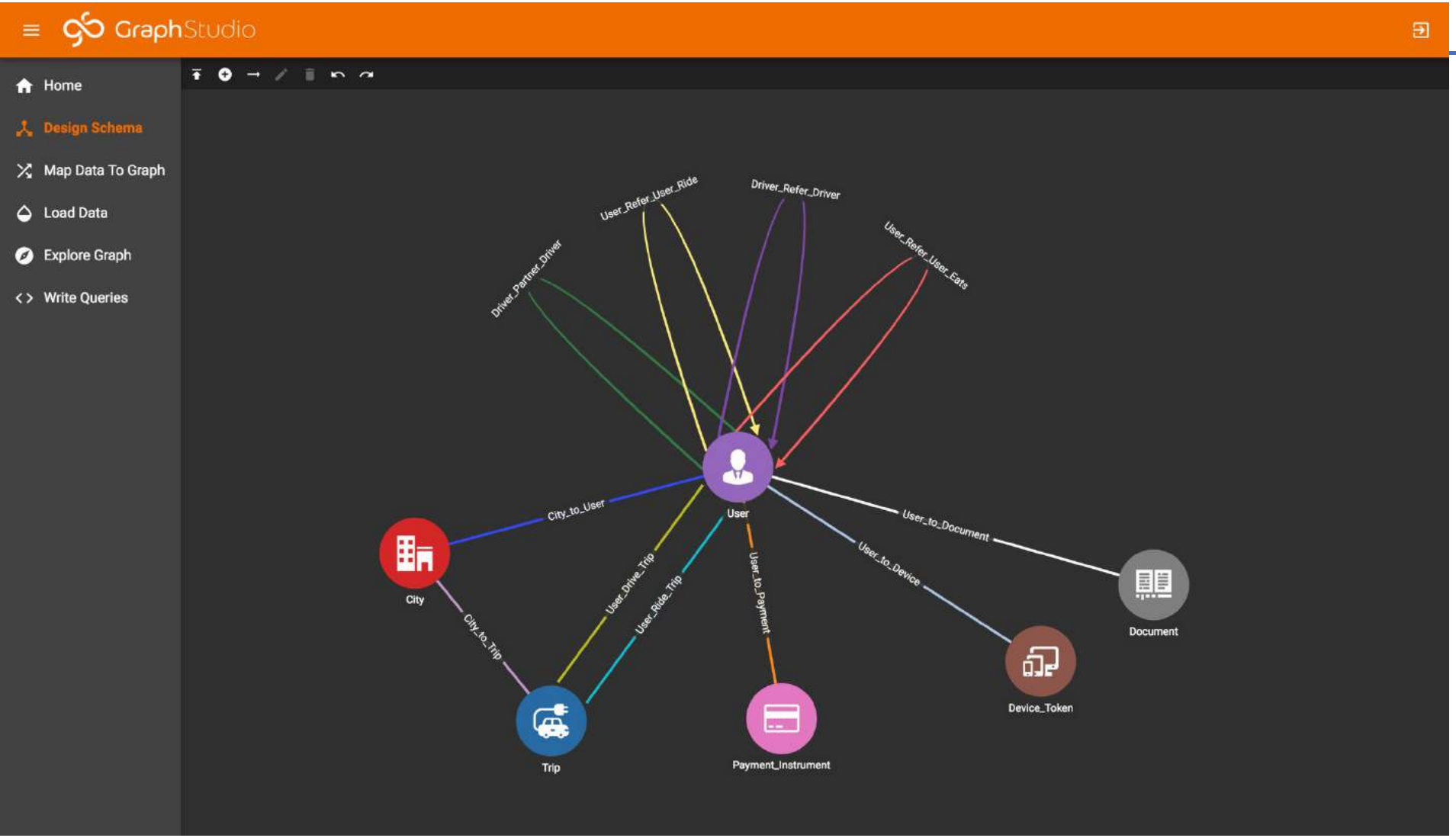
GraphStudio: Visual Graph Schema Designer



© 2017 TigerGraph. All Rights Reserved.

TigerGraph

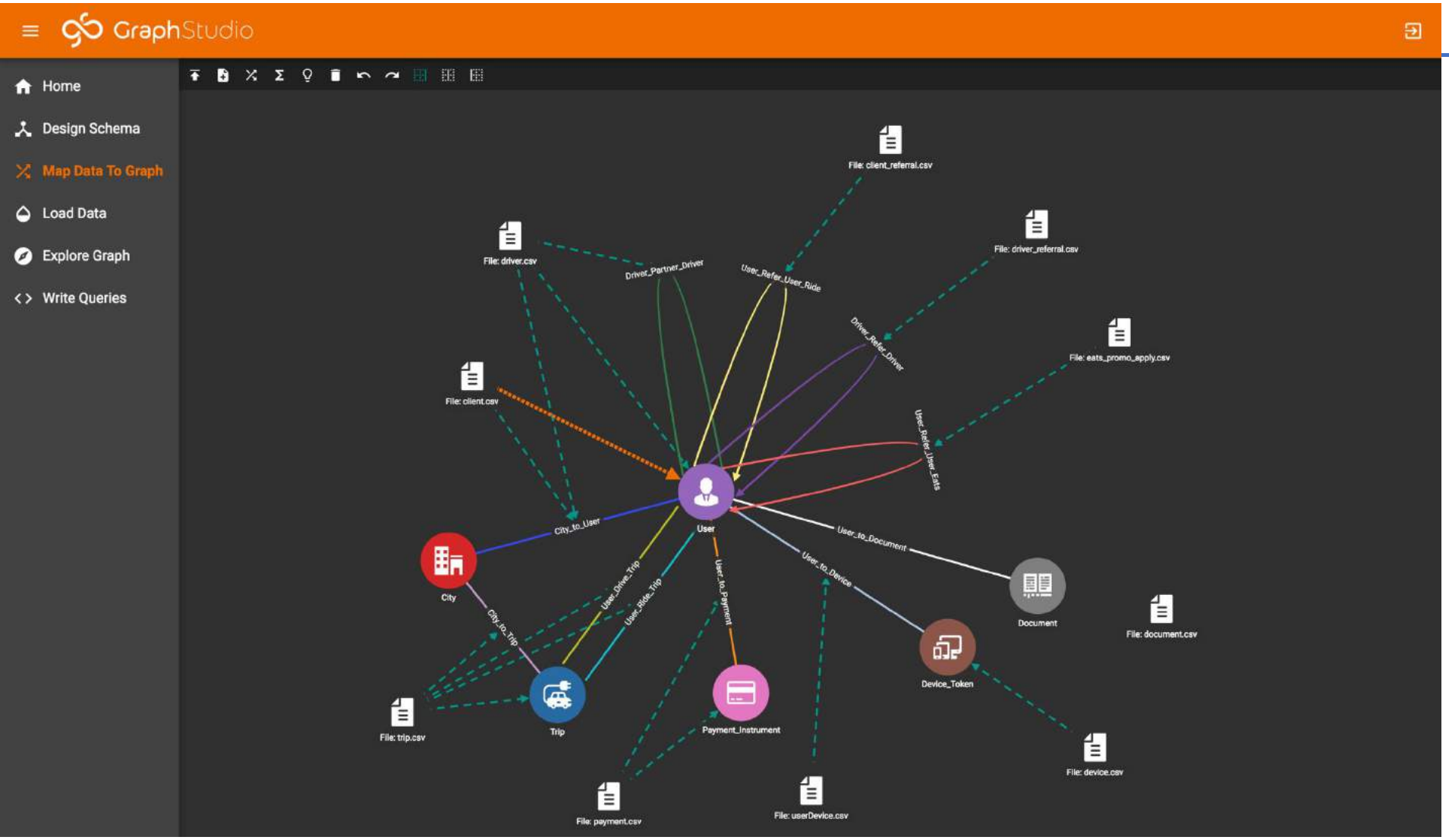
GraphStudio: Visual Graph Schema Designer



© 2017 TigerGraph. All Rights Reserved.

TigerGraph

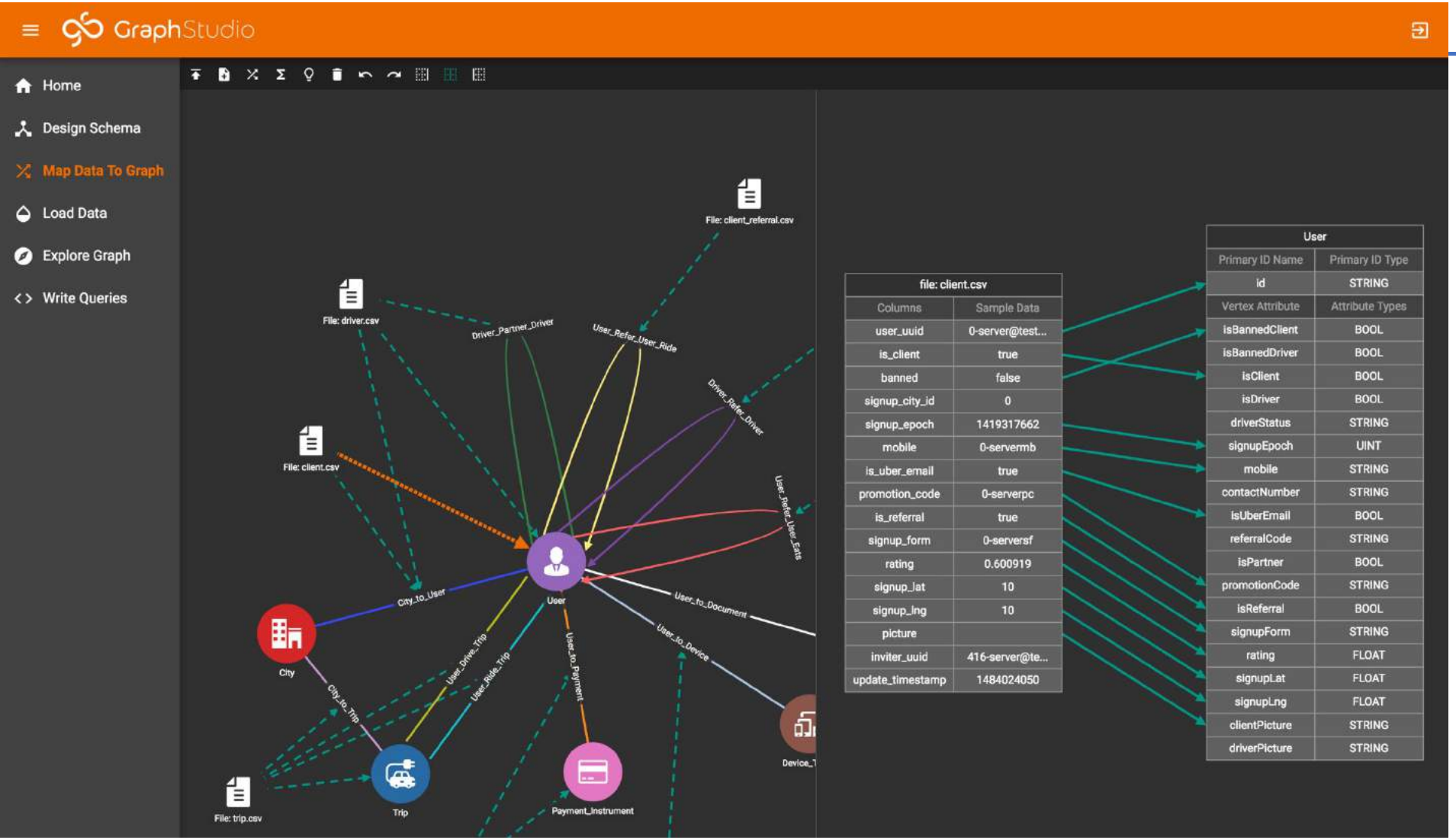
GraphStudio: Visual Graph Data Loader



© 2017 TigerGraph. All Rights Reserved.

TigerGraph

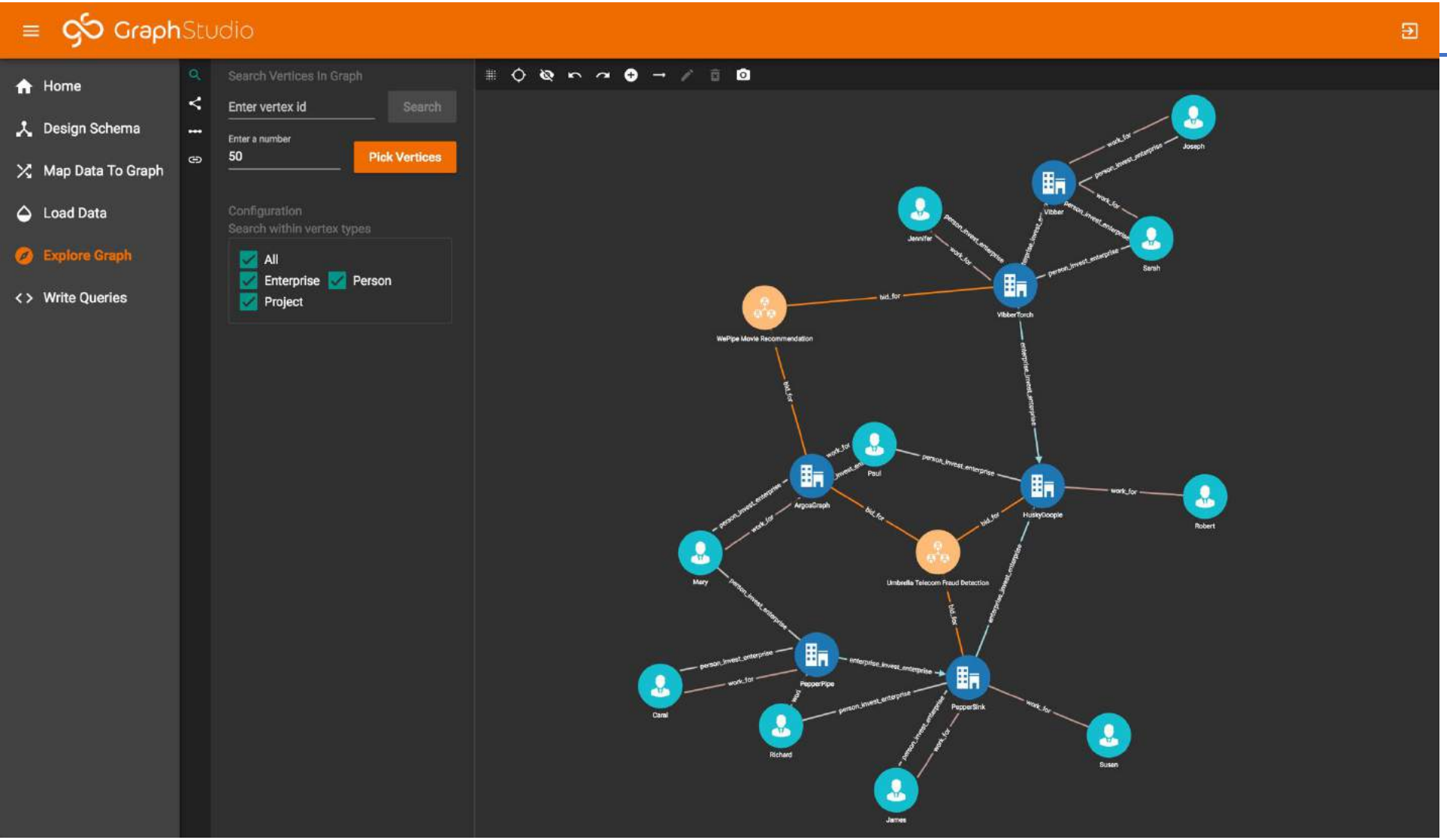
GraphStudio: Visual Graph Data Loader



© 2017 TigerGraph. All Rights Reserved.

TigerGraph

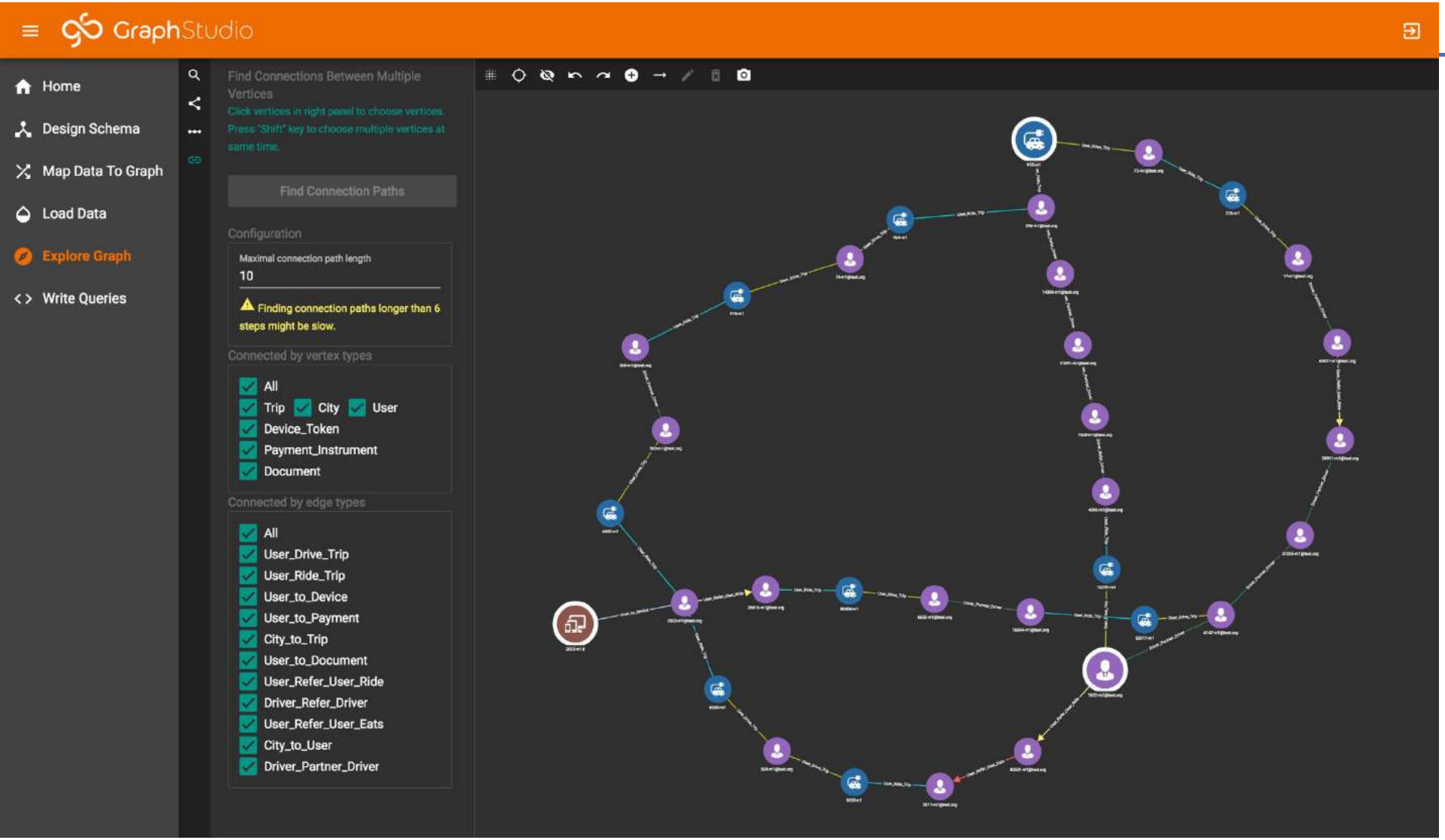
GraphStudio: Graph Explorer



© 2017 TigerGraph. All Rights Reserved.

TigerGraph

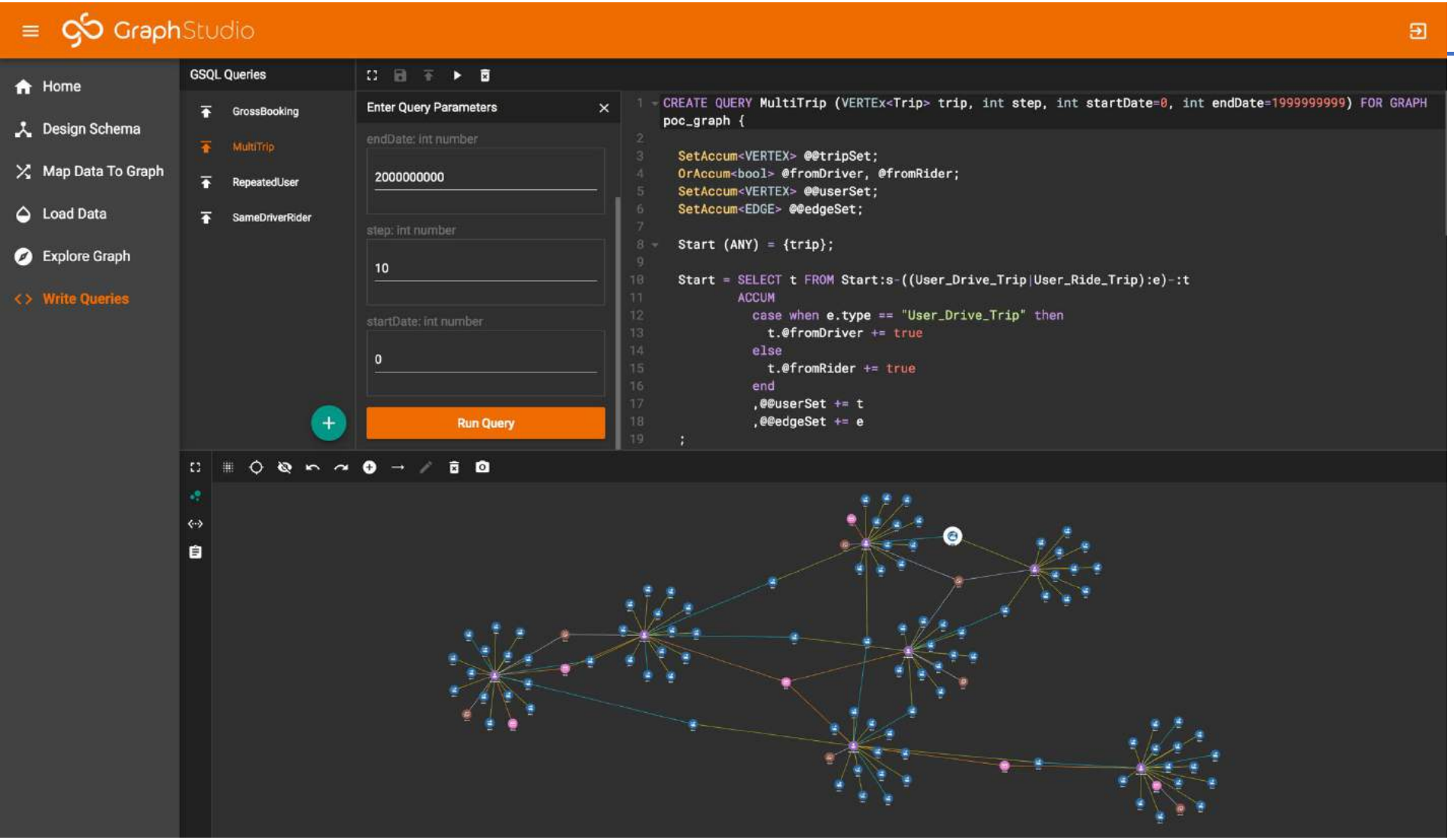
GraphStudio: Graph Explorer



© 2017 TigerGraph. All Rights Reserved.

TigerGraph

GraphStudio: Graph Query Editor



© 2017 TigerGraph. All Rights Reserved.

TigerGraph

TigerGraph is Hiring!

- Senior Technical Sales Engineer
- Senior Technical Product Manager

$3000 Referral Bonus! Contact us talent@tigergraph.com

See all our open positions at www.tigergraph.com/join-us/

© 2017 TigerGraph. All Rights Reserved.

Thank you!

TigerGraph

Follow us on Twitter or LinkedIn

to keep up with new developments and opportunities

More questions?

victor@tigergraph.com

info@tigergraph.com

© 2017 TigerGraph. All Rights Reserved.