Real-Time Deep-Link Analytics for Big Graphs

Challenges and Solutions
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.
The Data Deluge

Data Never Sleeps 2.0
from Domo.com

The Networked Age

Social Networks

Internet of Things

BIG DATA

© 2017 TigerGraph. All Rights Reserved.
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

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

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
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

Data Sources:
- CSV/Text
- Social
- RDBMS
- Hadoop Spark
- Log Files

TigerGraph Analytics Platform:
- Graph Query Language
- Graph Visualization
- REST API | Java | C++
- Standard UDFs
- Custom UDFs
- Graph Storage Engine (GSE)
- Graph Processing Engine (GPE)
- Graph Data Storage
- Graph Data Compression
- Parallel Processing
- Graph Partitioning

Enterprise Data Infrastructure:
- Business Intelligence
- Analytics
- Visualization
- Dashboards Reports
- Data Warehouses
- Master Data Stores

Infrastructure:
- On Premise
- Cloud
- Hybrid

© 2017 TigerGraph. All Rights Reserved.
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
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.
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.
A Closer Look At the TigerGraph System

GraphStudio

web browser based tool to visually and interactively define, create, explore and query big graphs
GraphStudio: Visual Graph Schema Designer
GraphStudio: Visual Graph Data Loader
GraphStudio: Graph Explorer
GraphStudio: Graph Explorer
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/
Thank you!

Follow us on Twitter or LinkedIn to keep up with new developments and opportunities

More questions?

victor@tigergraph.com
info@tigergraph.com