Google Vertex AI and TigerGraph: More Power for Analyzing Connected Data

Victor Lee, TigerGraph

Benazir Fateh, Google

Victor Lee Ph.D.

Vice President of ML / AI

TigerGraph

Product leader and educator, with a passion for algorithms, languages, user experience, and ethics. 7 years at TigerGraph, 3 years as university professor, 20+ years in tech industry.

Benazir Fateh Ph.D.

ML/AI Specialist

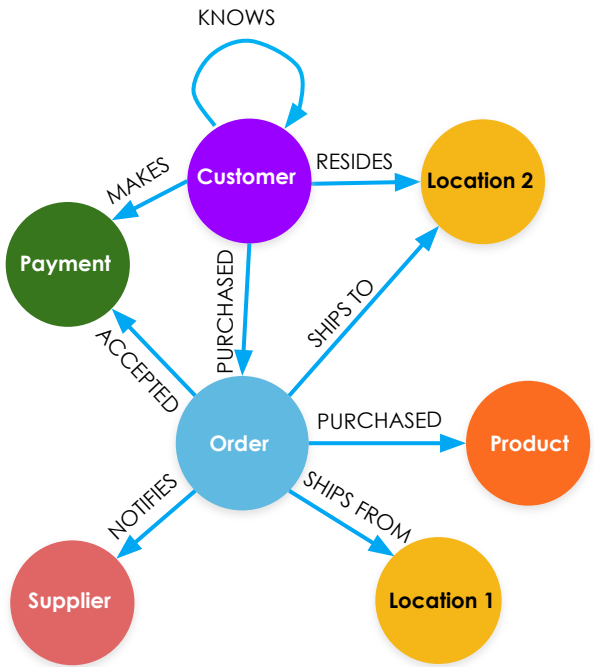
Google

AI/ML Specialist helping Google Cloud's largest customers create AI/ML solutions. Over an year at Google and before that 8 years in retail and precision agriculture.
Board member at an NGO helping immigrant youth.

Outline

- Why Graph + ML? (Victor)
- TigerGraph Cloud: scalable Graph Database for Analytics (Victor)
- Google Vertex: Unified ML Development and Ops Platform (Benazir)
- Vertex Pipelines (Benazir)

Why Graph? Why Graph + ML?



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

Three Basic Approaches for Graph+ML

1. Unsupervised Learning via Graph Algorithms
2. Feature Enrichment from Graph Feature
3. Graph Neural Networks

TigerGraph ML Pipeline: Data and Feature Prep

2 In-Database Enrichment

- graph algorithms
- graph pattern matching
- graph embedding

1 Build Your Graph

New external data
(IoT, web, streaming)

TigerGraph CLOUD

3 Options...

1. Export data (table or graph) to ML training platform

vertex.ai

Data from warehouses and other databases

2. Merge data (table) back to data warehouse or other database

3. (Coming soon) In-Database training

GNN

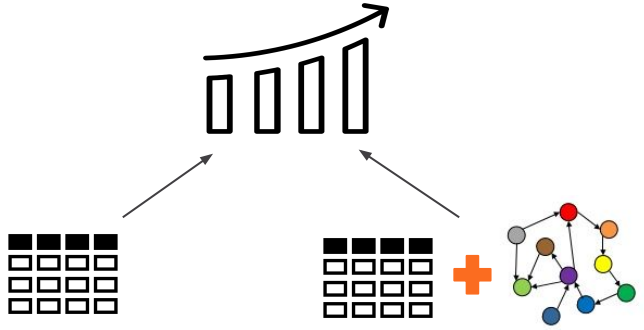
Using TigerGraph for Feature Engineering

In-Database Graph Algorithms

- 50+ algorithms for similarity, clustering, best paths, community, etc.
- No export needed. Live, updatable data
- Open-source, customizable
- Scaleable, Ultra-fast MPP graph engine
- Toolkit, building blocks for

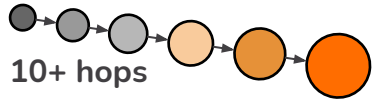
Graph Features Enriching ML Quality

- Graph enriches ML training data: quality in ⇒ quality out
- Use graph algorithms, pattern-matching queries, or graph embedding



Solving Enterprise-Scale Business Problems

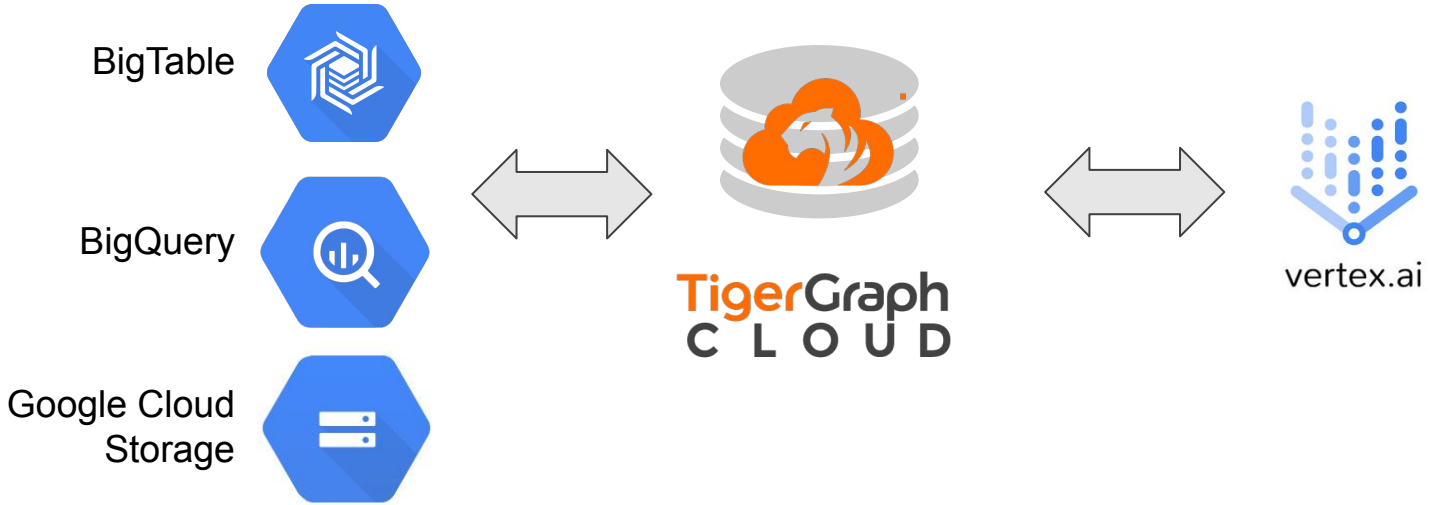
TigerGraph: Scalable Graph Database for the Cloud

Feature	Design Difference	Benefit
Real-Time Deep-Link Querying  5 to 10+ hops	• Native Parallel Graph design • C++ engine for high performance • Storage Architecture	• Uncovers hard-to-find patterns • Operational, real-time • HTAP: Transactions+Analytics
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

Connecting TigerGraph to other Cloud Services

Several ways to connect:

JDBC, REST APIs, pyTigerGraph Python driver, TigerGraph CLI



Summary

- Graphs provide **richer, smarter data** and **smarter, deeper analytics**
- Graphs have always had a **natural role in machine learning**:
  - Unsupervised learning through **graph algorithms**
  - **Graph features** provide richer training data
  - **Graph embeddings** transform graph data into easier-to-use vectors
- **TigerGraph Cloud** is designed for scalable and updatable graph analytics
- **Google Vertex AI** offers a **managed ML platform** for the cloud
- **Vertex Pipelines** simplifies MLOps

Thank you for watching!

- If you joined the live session, please use chat on the right side of the portal to ask questions. Speakers will be available to chat during the first session air date.
- If you have joined after the live session, please email marketing@tigergraph.com with any questions.
- To get started, go to
  - www.tigergraph.com/cloud
  - cloud.google.com/vertex-ai

Benazir:
How to do analytics with graph- not touching on analytics at all- only ML and Ops,

Victor:
Draw the connection between the evolution over time of high performance query, and graph algorithms, graph analytics and now bringing in machine learning into picture- TG is providing a richer source of data- with feature generation- if you do supervised learning, you need features to work with, and graph can provide additional features if you didn't have pattern matching. We identify features in data that can be piped into training system. Another way- you can use graph algorithms. More exploratory phase, data preparation, that can feed into traditional machine learning pipeline. We traditionally talk about analytics but we want to talk about machine learning and system design outside of TigerGraph.

Vertex- will talk about what Vertex platform allows us to do- different functionalities- Sagemaker demo- he dives into deeper part of tech- should we dive deeper into group of Vertex products. Useful to share with audience that such a breadth exists and touch on it, so audience understands Vertex AI, and spend a little time to think about more traditional simple pipeline but where automation aspects or model aspects are to highlight, why is it a step above.
Bruce: Don't cover all Vertex AI, but pipeline and data ingestion should be covered, metadata management, model registry, Pipeline is top priority but if that doesn't speak to Vertex's power

Pipeline is their star product that allows metadata management, lineage, will focus on pipelines and on one slide Vertex Pipeline means that you're already integrated with metadata and new feature coming out. Talk about prediction too, need to host this model on GCP product,
Want to record by end of next week but whenever Benazir and Victor have slides ready

Benazir first:

Victor second:
Victor will go through slides and put a TG icon and show how we fit in in the different stages. Extracted from TG and ingested in GCP and run on Vertex AI? Up to user
Google cloud storage or in another google warehouse, take data from multiple sources from TG and merge into one connected view- putting into graph, running feature graph extraction for data preparation, two options- straight into Vertex AI or go back into another database- graph will be used to figure out new features but the new features stored back into original database and train from there- where the model goes back- what was the last stop; some graph features and some others and where do you consider source of truth?

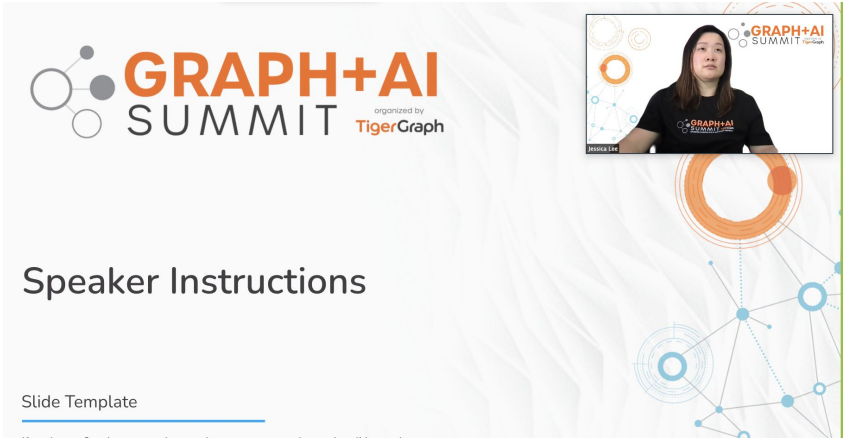
Instructions for Speakers

Question	Answer
When do I need to have my presentation ready?	Recording your session by September 30, 2021. Please have your presentation ready by your recording date.
Which slide template can I use?	You can use this Google Slide deck. To download it please make a copy of the Google Slide deck or download as a PowerPoint (Choose File → Download as Microsoft PowerPoint .pptx)
When will my session be recorded?	Please work with your TigerGraph counterpart to schedule and record your session. Recorded sessions can be uploaded here: https://drive.google.com/drive/folders/1Y7Z0RtiJoqPuGtHo_4oIKWObAvrl0IN_
How much content should I plan?	30 Minute Session: Plan for 20-25 min (max) of content for 30 min session (last 5 minutes are for live Q and A) 1 Hour Session: Plan for 45-50 min (max) of content for 1 hour session (last 10 minutes are for live Q and A)

Zoom Background Instructions

ZOOM BACKGROUND FOR SPEAKER VIDEO

- Please use the Graph + AI Summit Zoom background here.
- Go to Preferences under Zoom.us --> Backgrounds and Filter --> Click on the + sign to the right of "Virtual backgrounds and upload the file



Shared screen with active speaker

If you share your screen with the active speaker thumbnail, the recording will display the active speaker thumbnail in the top-right corner. If you drag the active speaker thumbnail to another location during the meeting, the recording will still display it in the top-right corner.

Recording Tips for Speakers

- Use a stand or a stack of books below your laptop to raise it up, so the camera is at eye level
- Lighting - Sit in a room that has ample natural light or lamps (lighting should be in front of you or above you, not behind you for best recorded video.)

Session Q & A Details

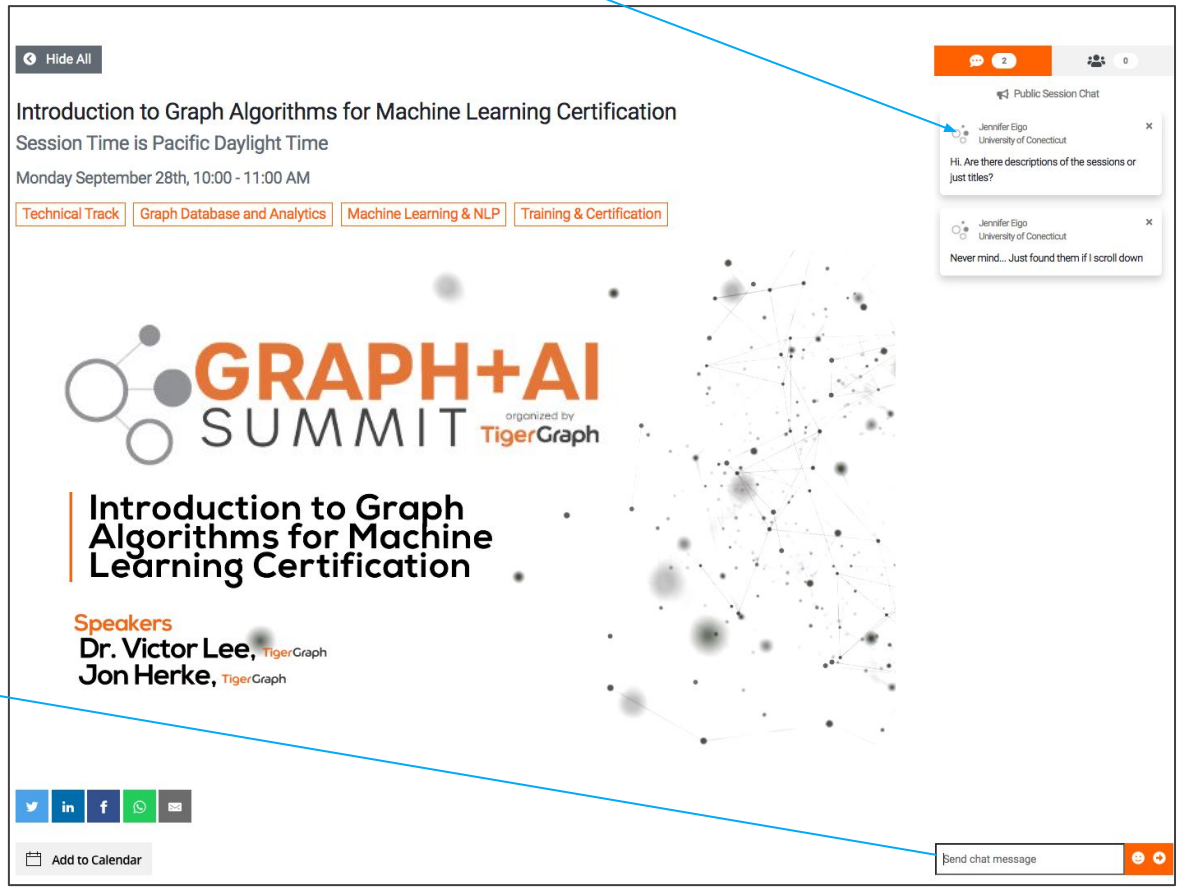
- Each session has 5 minutes of question and answer time at the end. For a 30 minute session, we are asking speakers to record 25 minutes of presentation, which will be streamed to attendees, followed by a live Q and A. **We are requesting speakers to join the session (if their schedule permits) and especially, the 5 minute Q and A at the end of the session.**
- In addition to the end of the session Q and A, attendees can enter their questions during the session in the Graph+AI Summit attendee virtual portal
- Speakers can join live during their session if their schedule permits and answer the question during the the session as well as at the end during 5 minute Q and A window
- If speakers aren't able to join for the live session due to timezone or schedule constraints, Graph + AI Summit team will be sure to get them the questions, so that speakers can answer those and we can share those back with the attendees

Q&A / Chat

NOTE: SESSION CHATS ARE PUBLIC

If you are available during your session time to be on Q&A via chat:

1. *Login to the virtual portal before your session. If you can't login to the portal please let us know at marketing@tigergraph.com*
2. *Go to your session at the time it airs (use search or filter to find your session)*
3. *You will be able to answer questions here - the chat will appear on the right side. All chats will be seen unless you send private messages (click on the person to send privately).*



Key Dates - Summary

- **By Thursday, September 30, 2021 5:00 pm US Pacific Time**

  Work with your TigerGraph counterpart to schedule and record your session. Please save your recording here: https://drive.google.com/drive/folders/1Y7Z0RtiJoqPuGtHo_4oIKWObAvrl0IN_

- **By Scheduled Recording Date**

  Have your presentation deck ready for recording

- **During October 4-19, 2021**

  If possible, please join your live session for Q&A with hundreds of attendees *(Please advise us during the session recording if you will be available for the entire session or just for the last 5 minutes of Q&A)*

Divider Slide