



The Power of TigerGraph:

# Seven Reasons Why TigerGraph is a Better Choice for Enterprise AI



**WHITE PAPER**

# Introduction

Analytics has become increasingly important for enterprises as they realize the value of the information that can be obtained from company and customer data. Now, it is just a matter of enterprises properly harnessing the right data and using it wisely to further improve internal operations or have a more granular understanding of customer data so they can offer more innovative, personalized services in real-time. As computing power becomes more robust and more affordable, artificial intelligence (AI) and more sophisticated data modeling such as machine learning (ML) are more readily available.

According to Gartner Research, in 2021, AI augmentation will create \$2.9 trillion of business value and 6.2 billion hours of worker productivity globally.<sup>1</sup> At the same time, McKinsey & Company reports that, from a recent survey, respondents coming from a variety of industries attribute 20 percent or more of their organizations' earnings before interest and taxes (EBIT) to AI, and these companies plan to invest even more in AI in response to the COVID-19 pandemic and its acceleration of all things digital.<sup>2</sup> To further validate the interest in AI, Forbes reports that 55% are on the AI journey with enterprises spending 62% more on AI than the previous year.<sup>3</sup>

## Data Pain Points

While enterprises report ongoing investments in AI, why are they struggling to capitalize on the potential of AI and machine learning? A significant reason is legacy data tools make achieving ROI in these areas difficult. Since most data is tabular and typically stored in two-dimensional table formats, this creates a gap in terms of how the data is interpreted in real-world scenarios. In other words, without the proper machine learning models, data interpretation can be lost or incorrect. This creates skepticism as recent research by Deloitte indicates that executives are worried about making the wrong decisions based on AI recommendations, and it is a top-three concern for 44 percent of respondents reporting.<sup>4</sup>

Another issue is data quantity and quality, especially as companies still have data silos across their organization. If data is distributed across non-integrated tools, it is difficult to use it in aggregate, meaning important strategic insights may be hidden beneath architectural complexity and ML is confined to operating in small pockets. SAP indicates that 74% of enterprises say their data landscape is so complex that it limits agility and 85% of enterprises struggle with data from a variety of locations.<sup>5</sup> Successful ML models thrive on both richness and volume of data. Without it, it's often not possible for AI to learn or produce useful insights.

---

“Complex business problems require contextual awareness and an understanding of the variable nature of connections and strengths across multiple entities, such as organizations, people or transactions. With graph technologies, critical business questions that used to take months to answer can now be resolved in minutes.”

**- Gartner**

---

<sup>1</sup> Gartner, August 5, 2019, Gartner Says AI Augmentation Will Create \$2.9 Trillion of Business Value in 2021 [Press release] <https://www.gartner.com/en/newsroom/press-releases/2019-08-05-gartner-says-ai-augmentation-will-create-2point9-trillion-of-business-value-in-2021>

<sup>2</sup> McKinsey & Company. The State of AI in 2020. November 17, 2020. <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2020>

<sup>3</sup> Louis Columbus, “Enterprises Increased AI Spending By 62% Last Year,” Forbes, March 15, 2020. <https://www.forbes.com/sites/louiscolombus/2020/03/15/enterprises-increased-ai-spending-by-62-last-year/?sh=6eb39d1a1ead>

<sup>4</sup> Susanne Hupfer, Deloitte Insights, Capitalizing On the Promise Of Artificial Intelligence: Perspectives on AI Adoption From Around the World, 2019. [https://www2.deloitte.com/content/dam/insights/us/articles/6458\\_capitalizing-on-the-promise-of-ai/DI\\_Capitalizing-on-the-promise-of-AI.pdf](https://www2.deloitte.com/content/dam/insights/us/articles/6458_capitalizing-on-the-promise-of-ai/DI_Capitalizing-on-the-promise-of-AI.pdf)

<sup>5</sup> SAP, Data 2020: State of Big Data Study, A Global Study of Enterprise Data Challenges and Solutions, <https://www.sap.com/documents/2019/06/4c0526ad-537d-0010-87a3-c30de2ffd8ff.html>

Furthermore, some enterprises fail to operationalize data and AI ethics. With privacy and data regulations in both Europe and the U.S. being updated on a regular basis, not being up-to-date in these areas can put companies at regulatory and legal risk. Not just the legal ramifications, but lack of operationalized data makes it difficult for systems or employees to know whether or not to take action if there is not a clear recommendation from an AI/ML engine. ML results are often derived through highly statistical, deeply complex and transparent processes, which humans are not able to keep up with manually, so enterprises need data tools that have modern AI and ML modeling.

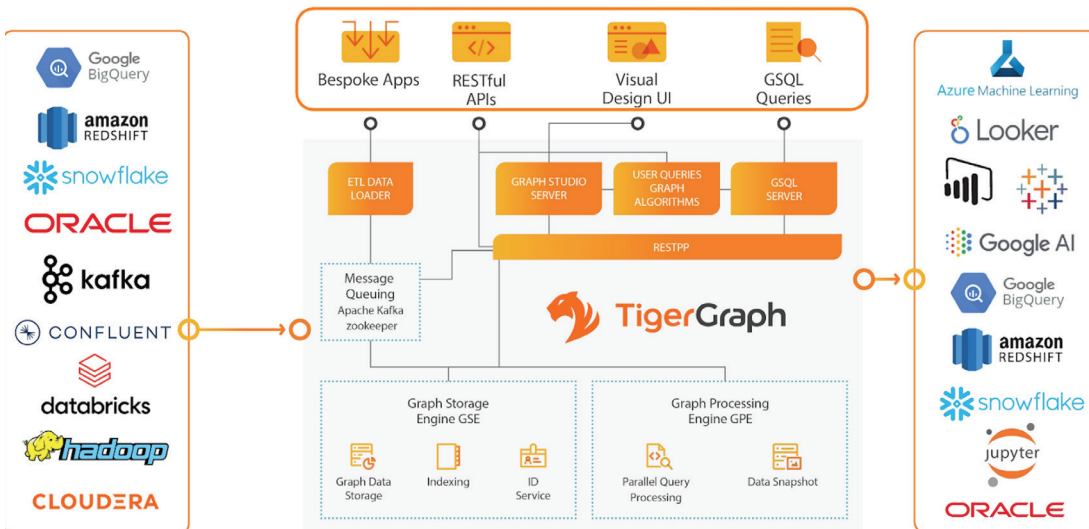
## Tabular Data Becomes Connected Data

With deep roots in data management and optimization, TigerGraph helps enterprises achieve the potential of analytics and machine learning by connecting data at scale. TigerGraph connects enterprises' data in the form of real-world relationships, in exactly the way humans comprehend information. This means machine learning models using graph data input can process information in a human way, using both data points and the relationships between them as features. This generates results that make a lot more sense to those interpreting the data.

As TigerGraph connects together datasets from disparate data sources, this enables enterprises to easily search as broadly and deeply as needed to uncover strategically important insight. Connecting data addresses the bad data quality and the issue of missing data points. As ML models thrive on high volumes of great quality data, they can also thrive on low volume data where there is relevant context (i.e. connections). In other words, having the right quantity and quality of data is now feasible in any data volume scenario.

Explainability and precision have now become the norm. GraphStudio visualizes results and enables interrogation in such a way that we can comprehend the reason for an AI decision or recommendation—making AI more explainable. TigerGraph has created a culture of data evolution where connected data maintains a fluidity, new data and connections are made continuously, and ML models move towards the truth. TigerGraph completes a complementary ecosystem with an enterprise's data lake, data visualization and AI tools.

**FIGURE 1. TIGERGRAPH ECOSYSTEM**



## How TigerGraph Unleashes AI Potential

With its speed, scale and analytical sophistication, TigerGraph has created a new genre of native graph database that enables the full potential of AI to be realized. This is comprised of three key attributes:

**Low Effort Scaling** - Connecting all of the data across your organization is going to require a graph database that handles scale.

- Time it takes to get to the answer
- Readiness for enterprise scale
- Return on investment
- Transparency during calculations

**Advanced Analytics** - Revealing all strategic insight within your data will require top-tier graph analytical capability.

- Breadth and depth of questions supported
- Variety of questions it is possible to ask

**Organic Insight** - Evolution of real world connected data underpins ML's journey to the truth

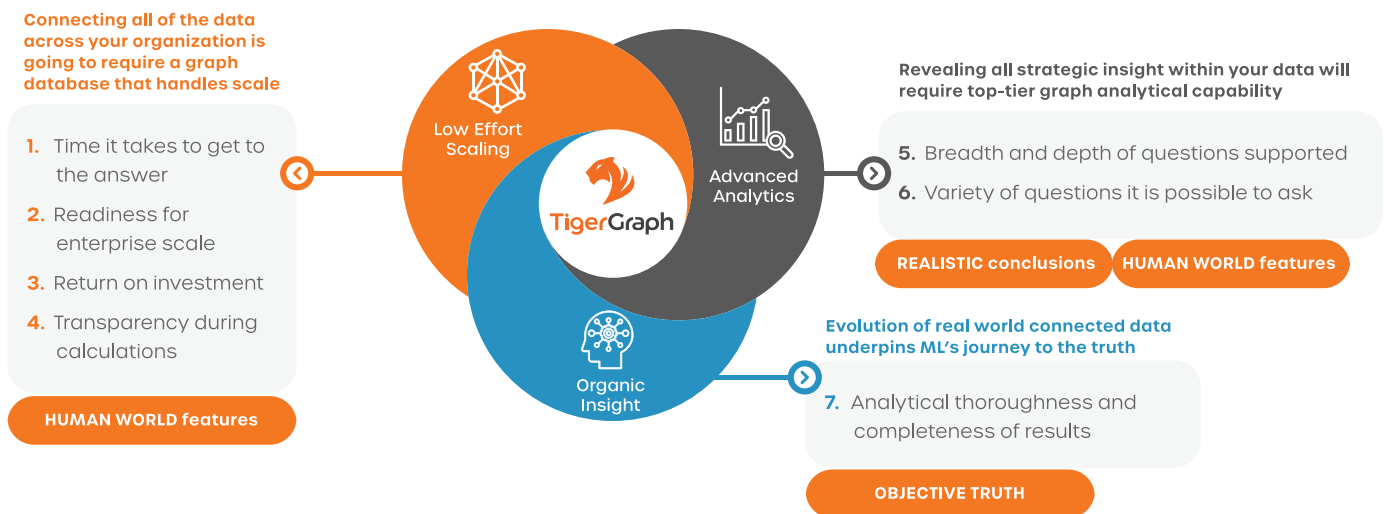
- Analytical thoroughness and completeness of results

With its speed, scale and analytical sophistication, TigerGraph has created a new genre of native graph database, built to help realize AI's potential.

“TigerGraph’s ability to handle large quantities of data coupled with their elegant and powerful query language GSQL have enabled us to build a graph data warehouse which we use to help our users understand flows of funds and determine their risk exposure. [TigerGraph has proven to be invaluable in helping our users to differentiate between good actors and bad ones.](#)”

**- Nirmal Aryath Koroth,  
Co-Founder and Chief  
Technology Officer,  
Merkle Science**

**FIGURE 2. TIGERGRAPH NEW GENRE OF NATIVE GRAPH DATABASE**



# Seven Reasons to Choose TigerGraph

TigerGraph's unique capabilities to connect, analyze and provide insight at scale, deliver these seven benefits:

## 1. Speed of Query Results

When it comes to query results, the question is often asked, when and how is speed required?

Because a graph database connects data in a way that has never been possible before, it is creating the possibility to ask important new types of business questions.

Business questions today may be highly complex because they relate to important and strategic cross-functional business priorities. Today's business complexities have never been computationally supportable, because computing processes to connect data have been too time-consuming or resource-intensive. Another barrier is the insights themselves may be so unique it requires some human live intervention in order to understand data and take action. There may also be questions of critical financial importance and can, for example, result in immediate, realtime action that prevents financial crime.

Why is speed so important? Even for simple queries, a high speed graph database runs much more efficiently. High speed is also non-negotiable in the following scenarios:

- **In the case of very complex queries, especially where the data is changing or growing in the background**
- **If you are using graph for data investigation where speed is imperative**
- **If you are driving an operational process that requires realtime or near-realtime information, and speed of results is paramount**

As data is accumulating exponentially across the enterprise, graphs gain traction over time, and speed is only going to become more important, and even waiting an hour or two, or even overnight just to ask the next question of your data, can be unproductive and detrimental.

### How TigerGraph Compares to Neo4j

TigerGraph is able to search across more than 10 million data types and data connections per second per machine (the population of Greece) every second, and each additional parallel machine doubles this quantity.

TigerGraph can also support 100,000 updates to any data point anywhere in the graph every second, per machine.

Neo4j by contrast is 10 to 1000 times slower, as proven by [independent tests](#).

---

“With a medium-sized graph (82.7 million vertices/business objects, 218 million relationships) **TigerGraph finished four-hop queries in 30 minutes. However, with Neo4j we couldn't get those four-hop queries to finish after 10 hours.**”

**- Hiep Doan,  
Software Engineer, Gojek**

## 2. Readiness for Enterprise Scale

The purpose of a graph database is to connect data together to create new insights and information. Not only do graph databases have storage capability like a data lake but also have data processing engines. Any single database server ultimately has limited processing and storage power. In order to process very high volumes of data points at very high speeds, graph databases scale into additional servers.

Once a graph database sits across more than one server, there must be a level of consistency across the servers that enables the graph to operate as a whole when connecting data distributed across the servers and enabling users to ask questions that relate to all of the distributed data. This consistency is usually driven by a consistent data categorization model and an organized distribution of the data into each server, with the ability for your graph to operate seamlessly across all servers is important.

If users are required to manually distribute, load and query data across servers, not only is this inefficient, but processes may be plagued by human error, which can mean the difference between right or wrong insights. Automatic scaling across servers ensures the integrity of a data query, and automation also eliminates the need for manual rework of categorizations and distribution, as well as on the queries that have been set up already.

### How TigerGraph Compares to Neo4j

Both TigerGraph and Neo4j scale out to process higher volumes of data using more servers, but there are key differences in terms of automation and flexibility.

With Neo4j, the user must manually do the following for each server:

- **Set up a new data categorization model for each new machine**
- **Divide the data and load separately into each machine**
- **Write queries in such a way that they reference the right data stored in the right machine**

The manual nature of these steps can lead to human error—and ultimately cause inaccurate results. Sometimes, there simply is no sensible way to do manual splits.

TigerGraph always automates all data distribution, categorization across servers and cross-server processing of queries, without the user having to do anything manual after provisioning.

## 3. Return on Investment

Return on Investment (ROI) depends on how much you invest and how much value you derive over time. TigerGraph is not just powerful, but also efficient handling the same amount of data and querying with fewer machine resources than its competitors. This maximizes initial ROI. Moreover, data and use cases tend to increase over time, so the ability for a graph database to scale out and to handle new applications is vital for longer ROI. TigerGraph's unlimited expansion without the need to rearchitect the data model, along with the much-praised power of its GSQL query language, mean that your investment is safe with TigerGraph.

---

**“Graph algorithms scale exponentially.** Graph requires scalable software, more so than any of the other situations or challenges you have considered.”

**- Brad Spiers, Executive Director, JPMorgan Chase**

---

“We were very clear that we needed to find a graph database that partitioned across a distributed network. Neo4j was easy to get hold of and play around with. For small models, it's great; for models that are relatively homogenous, with not too many links, it does fine. **What we found was, as soon as we move beyond [a] point model, we just hit a wall.**”

**- Harry Powell, Director of Data & Analytics, Jaguar Land Rover**

## Return on Investment – Cost

**FIGURE 3. WHAT DRIVES THE COST OF USING A GRAPH DB?**

Cost	TigerGraph	Neo4j
Graph DB is priced by:	...minutes of processing, timed by RAM	...minutes of processing, timed by RAM
Infrastructure cost	Storage, Computes	Storage, Computes
Employee time cost	Query creation, Query processing	Query creation, Query processing
Loading 967 MB of data results in infrastructure costs for	482 MB, or <b>50% of the size</b> - due to TigerGraph compression	2300 MB or <b>237% of the size</b> - due to Neo4j indexing expansion

The higher speed the database, the less it costs to use.

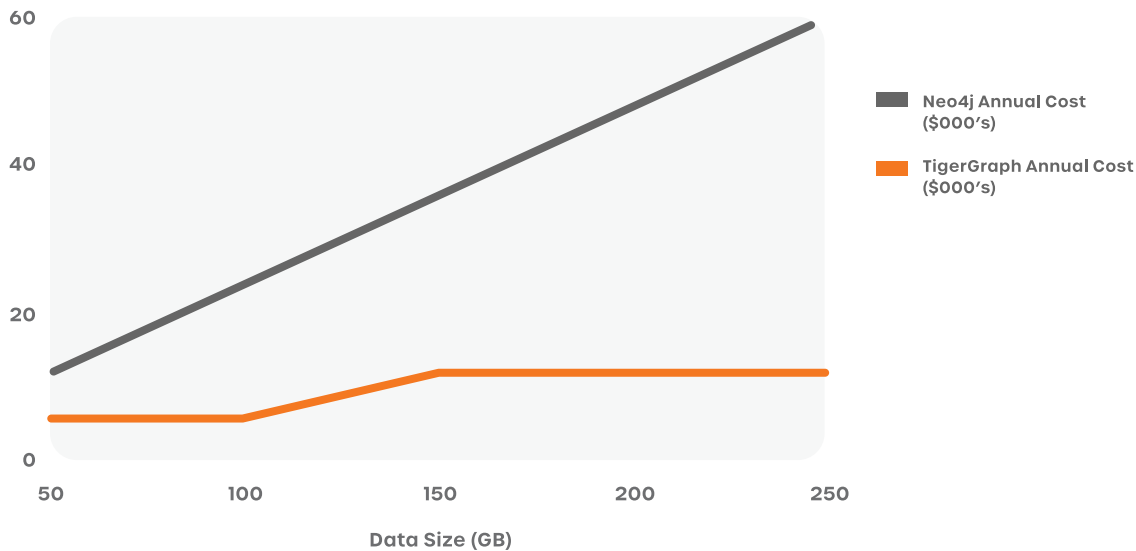
**TigerGraph is up to 1000 times faster than Neo4j.**

The more efficiently the graph database stores and processes data, the less it costs to use. **TigerGraph compresses your data, whereas Neo4j's indexing process expands it.**

The **easier the graph query language, and the faster the graph processing speed**, the less employee time required to use it. TigerGraph's GSQL is so similar to SQL it's very simple for users to learn quickly.

The following figure shows the annual computing or hardware costs for TigerGraph vs Neo4j based on the memory or RAM requirements for each graph database. As illustrated, the cost of TigerGraph is 50% lower than Neo4j. To learn more about TigerGraph, you can download the [TigerGraph Buyer's Guide](#).

**FIGURE 4. COMPARISON OF ANNUAL HARDWARE COSTS**



## 4. Transparency During Calculations

A data scientist can build the same graph query in a number of different ways, but some will be more efficient than others. Efficiency of graph query drives cost and user experience, and depending upon the way the query is written, the formation will ideally return results the fastest, and using the lowest processing power. This is why monitoring query performance is important as it gives users an understanding of whether to optimize them. With each amendment a user makes, it's important to be able to see what effect it is having on the query's performance.

## How TigerGraph Compares to Neo4j

TigerGraph's Administrator Portal allows users to visually monitor performance in realtime in a very visual and interpretable way.

With the stats provided on query performance, users can optimize how queries run—re-writing where it is clear there are efficiencies to be gained. Example stats include number of queries completed per second, percentage of queries that timed out and didn't finish, and minimum/maximum/average time to complete a query.

With the stats provided on how the overall machine is operating, users can manage storage and processing requirements. For example, users can see usage stats for CPU, Memory and Disk.

While there are some third party apps that have been created for Neo4j users, Neo4j does not have a native visual monitoring capability that users can use to optimize queries and performance. Instead it requires users to run command line functions to return results in a less format that is not easily interpretable.

## 5. Breadth and Depth of Questions Supported

The unique value of a graph database is that it relates data points together into a network and can bring out the context and meaning behind data. Now, users can discover chains of events in connected data, or might find relationships they didn't know existed, or patterns of behavior, or a groups of traits that prove important aggregates that might otherwise be siloed. As a graph database loads your data into a network and then allows users to ask questions of that data once it's in network form, this means you can ask a whole new set of questions. This highlights why functionality is so important. A comprehensive data tool gives users the power to ask any question they have about their data.

TigerGraph can support any number of connected data points—and any number of chained connected—in its queries. This is a game changer because without the power to analyze any number of connected data points together and to any depth, users will be limited to only searching part of their data network for the answer. In other words, it would allow visibility into only part of a causal chain, or only part of a behavior pattern. This means the possibility of missing out on crucial insights that might require immediate action—a missed opportunity.

More data also means more potential questions and insights, and the more granular data can be, the higher the integrity of data insights. For example, detailing your 'web' and 'in store' customer journeys respectively as 'on chat', 'on home page', 'on product page', and 'in city store', 'in suburban store'. This granularity means the database has to evaluate more possibilities, but the payoff for this added breadth is more context for finer grained insight.

Let's examine the breadth and depth required in two very simple example questions:

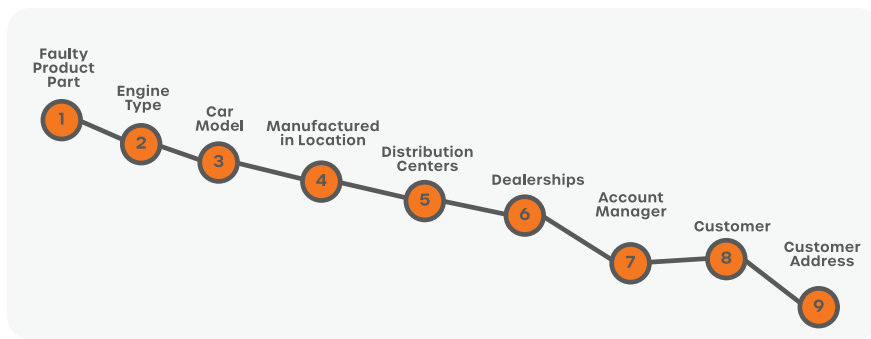
---

"We reached out to TigerGraph as we've been trying to find an elegant way to visualize our investigative data over the last 2 years. Other incumbents in the graph database space weren't able to process our vast amounts of data fast enough in order to generate graphs in real time. [TigerGraph helped solve that issue for us—with its ETL ingestion speed, we could do both batch and streaming load at the same time.](#) TigerGraph's GSQL software program is also the most sophisticated query language I've seen so far and it's flexibility allows us to implement complex graph algorithms which would otherwise be impossible or take far longer to implement on other incumbents."

**- Nirmal Aryath Koroth,  
Co-Founder and Chief  
Technology Officer,  
Merkle Science**



FIGURE 4. WHERE DID THIS FAULTY PRODUCT PART END UP?



In this query, the product part can be connected back to all the customer addresses at which it can now be found. This diagram simplifies the number of data types needed to answer this question. For example, a supply chain may have any number of categorizations to the left of 'car model', and there would likely be more channels of sale than just dealerships too

FIGURE 5. WHY ARE CUSTOMERS RETURNING PRODUCT X?



This query reveals that customers are returning Product X because if they buy it again now, they buy it cheaper. Note that the query did not ask about the price change, it asked what the reason was. So, the graph needs to be looking for all other potential reasons. Maybe a membership of a social media group? Or recent visits to a competitor product's website? That's possibly more connections for the query to analyze.

### How TigerGraph Compares to Neo4j

Neo4j supports asking questions across fewer than 5 connected datapoints, whereas TigerGraph alone can support any number of connected datapoints in its queries.

With Neo4j, Spark must be used (outside of the graph database) if queries require a higher number of connected datapoints, which means exporting data into Spark and running lengthy processes that aim to mimic graph capability. Users then have to reload results back into their graph.

This Spark process can take a lot of time and require additional loading steps for users. To make matters worse, its calculation process can actually time-out for questions that involve too many connections.

"In a globalized world, ever more complex corporate and financial structures are created for a variety of reasons, including fraud, money laundering, tax evasion, sanctions evasion, and more. [TigerGraph's graph analytics platform allowed us to go six to nine levels deep into the dataset - in real-time](#) - to uncover questionable patterns and potentially suspicious activities. These extra levels of 'data depth' are the difference between detecting all kinds of financial crime—or letting them go undiscovered."

**- Andrew Hoagland,  
VP of Engineering,  
Sayari Labs**

## 6. Variety of Questions it is Possible to Ask

As the breadth and depth of questions has been discussed, the next logical discussion would be the variety of questions that can be asked. Using a graph database means there is flexibility in the types of data you can load as well as the variety of questions you can ask of your data. In other words, a graph database addresses any business problem or use case that needs more insight from connected data. What's key, of course, is a data tool that gives users the flexibility to ask any question to the insights they want. This is particularly crucial in scenarios like investigative analysis.

The upside for enterprises is that graph databases come pre-configured with a standard set of questions ("graph algorithms") to ask of connected data. However, sometimes these general purpose tools are not enough to extract unique insights. For some enterprises, it may be necessary to ask more specific questions. For example, a question may have sub-contexts where a user may need to exclude or include certain types of data, based on a variable. Like: "Show me the customers that make referrals frequently, but only if they've visited our website in the last 24 hours." What's key in this scenario is the ability to fully customize questions. The ability to customize questions is also important when you are using a graph to enhance machine learning models.

TigerGraph provides users with the ability to design any question entirely from scratch or edit preconfigured questions. The company is able to do this because its standard algorithms are open source and its query language, GSQL, is a fully flexible, complete programming language. This means there is limited upskill required for data scientists who usually know SQL. This is in contrast to a standard-first approach, meaning users are not able to directly customize algorithms themselves and they are closed source.

### How TigerGraph Compares to Neo4j

In addition to the graph algorithms that come as standard with Graph Studio, TigerGraph also provides users with the ability to design any question (i.e. create any graph search algorithm) entirely from scratch, or by editing any standard question.

TigerGraph is able to do this because its standard algorithms are open source and its query language, GSQL, is a fully flexible, complete programming language. There is minimal upskill required for data scientists who almost always already know SQL.

By contrast, Neo4j supports algorithms only as fixed built-in functions, meaning that users are not able to directly customize algorithms themselves and they are closed source. Neo4j's query language, Cypher, is also intentionally limited, and Neo4j developers do all programming/customization on behalf of their users.

---

"TigerGraph's [GraphStudio offers a lot more functionality when it comes to explorations](#) and query management than Neo4j's desktop."

**- Hiep Doan,  
Software Engineer, Gojek**

## 7. Analytical Thoroughness and Completeness of Results

When it comes to data science, building data categories is building knowledge. As humans take their life experiences to identify and categorize things, graph databases categorize and relate data points together. For example, humans learn something new about a category, store the information in one's memory and remember to make any amendments the next time there is a query. Ideally, graphs mimic this categorization approach and organically amend and add categories as appropriate over time.

Why is categorization so important? When a data scientist asks a question of graph data, the graph needs to know how to answer the question. The more categories, the better likelihood a query will lead to an answer. If a data scientist can guarantee that every data point in a graph is categorized, then it can be guaranteed that every data point will be included in any question searching its category. Consequently, a graph without any data categorization, cannot guarantee that every data point has been assigned a category, and the data's integrity is in question as it may be inconsistent throughout.

There are two ways to load data into a graph database. The first is a category approach which loads each data point into a specifically defined data category, with specifically defined connections to other data categories. The second is a flat approach which is just loading it with the intention of maybe categorizing and connecting it later. TigerGraph makes it mandatory to take the category approach, whereas other vendors may be built with the flat approach as standard, and the category approach is only an option. Having only partial or no categories could lead to partial searches or incomplete results.

It almost never makes sense to load data without defined categories because without categories, you cannot ask the questions you need to ask of your data. It's important to remember that:

- **The primary purpose of a graph is to connect data and process it.**
- **If you don't categorize the data upon load into the graph, you run the risk of not categorizing it ever.**
- **Any data that is not categorized will not be returned in your query results.**
- **Graph databases are intended for use with structured or semi-structured data—the very need for structuring is in order to support graph algorithms, which draw on categories in order to provide answers to your questions.**

### How TigerGraph Compares to Neo4j

TigerGraph loads data according to the Category Approach, whereas Neo4j is built with the Flat Approach as standard, and the Category Approach enabled as an option.

TigerGraph maintains flexibility for users by enabling new Categories to be created or updated at any time.

Recently, Neo4j has started advising customers to opt for the Category Approach, but it remains a customer's choice as to which approach to take, meaning they could start with either approach, and choose to change approach at a later point in time.

---

**"TigerGraph reached up to 1.5 million created nodes per second, while for Neo4j that number stopped at 22,000 nodes per second and degraded over time."** We tried to load data with Neo4j in parallel by dividing our dataset into multiple files and tried to run multiple processes to load the data. We ran into problems with the locking mechanism in Neo4j where for those processes to run in parallel, they ran into lock contention problems and eventually, the loading could not finish."

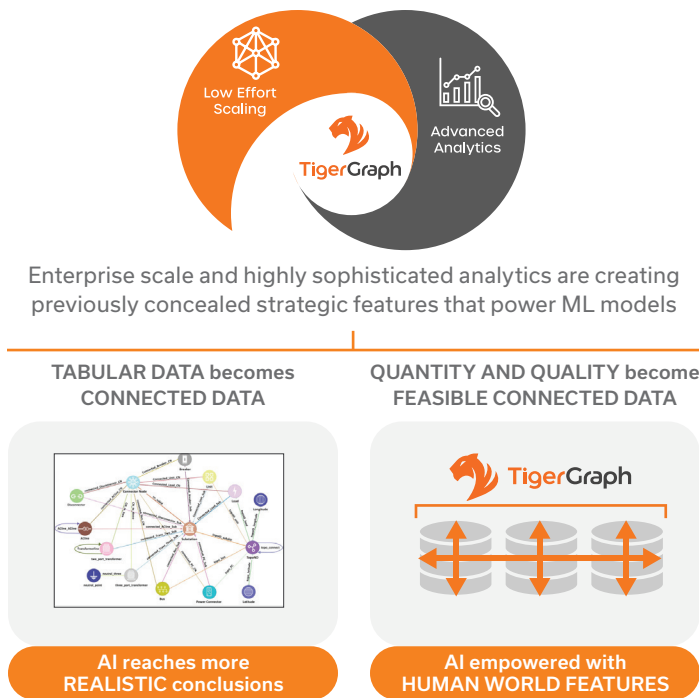
**- Hiep Doan,  
Software Engineer,  
Gojek**

## TigerGraph Allows a Data Graph to Evolve Organically and Create New Insights

Using a data tool like TigerGraph can allow a graph to evolve organically, which creates new insights that become data points in themselves for external AI and learning systems.

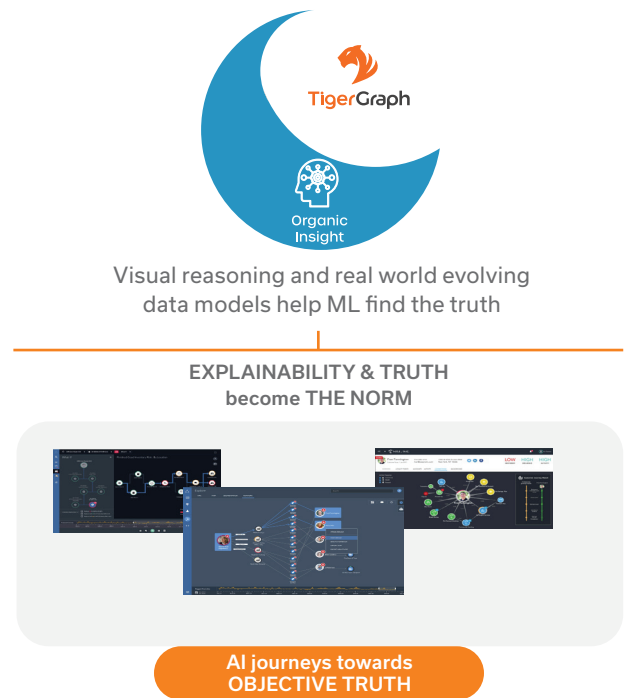
This means that categories always remain flexible for updates, and connected datasets will always reflect the evolving real world scenarios of an enterprise. This is important for ML models that learn over time. TigerGraph's mandatory categorization means an enterprise's data is consistently categorized, and there is no risk of extracting incorrect insights or machine learning features from incomplete data.

**FIGURE 6. ENTERPRISES REALIZE THE POTENTIAL OF AI AS THE NEW ORGANIZATIONAL BRAIN**



"As our work continues and our data grows, we had challenges scaling our data to meet our business needs. TigerGraph's excellent scalability and performance enables us to achieve things we previously could not do, and to better support ongoing investigative work in the process."

**- Chris Taggart, CEO, OpenCorporates**



## Conclusions

AI and ML have become increasingly important for enterprises as they realize they must be able to have a granular view of business and customer data in order to keep up with the competition. Now, it is just a matter of enterprises capturing and being able to analyze the right data and use it to take actionable insight in real-time and provide more personable products and services. As legacy data tools have left data in tab form, more sophisticated AI and ML tools such as TigerGraph allow for graph insights which show data points in their network and connected form, revealing aggregates, patterns and behaviors that have never been seen before. TigerGraph is flexible and able to scale with enterprises while allowing users to ask any question in order to find out what is needed for their specific needs.

### **About TigerGraph**

TigerGraph is the only scalable graph database for the enterprise. Based on the industry's first Native and ParallelGraph technology, TigerGraph unleashes the power of interconnected data, offering organizations deeper insights and better outcomes. TigerGraph fulfills the true promise and benefits of the graph platform by tackling the toughest data challenges in real time, no matter how large or complex the dataset. TigerGraph's proven technology supports applications such as fraud detection, customer 360, MDM, IoT, AI and machine learning to make sense of ever changing big data, and is used by customers including Amgen, China Mobile, Intuit, Wish and Zillow.

**The company is headquartered in Redwood City, California, USA.**

**Follow TigerGraph on Twitter at [@TigerGraphDB](#) or visit [www.tigergraph.com](http://www.tigergraph.com)**

© TigerGraph, Inc. 2021 All Rights Reserved.

**WP-Seven Reasons Why TigerGraph is a Better Choice for Enterprise AI-07062021**