HOW GenAI + HPC ACCELERATE GRAPH ANALYTICS ADOPTION

EXECUTIVE SUMMARY

The interconnected, tech-centric environment in which businesses and government organizations now operate has generated a long list of challenging insight discovery problems, from cybersecurity to advanced forecasting. Emerging AI applications—including agentic workflows that must synthesize diverse data points to perform reasoning and take action—are only growing in demand. Graphs are an effective way to represent the complex systems that underlie much of this new insight discovery. Now is the time to adopt and expand your organization's capabilities for **graph analytics**.

To date, two primary barriers have limited the widespread adoption of graph analytics:

- The need for specialized technical expertise to perform analyses
- Performance constraints related to handling complex, large-scale data queries.

Embedded generative AI (GenAI) addresses the first barrier by enabling your analysts to interact naturally with data without extensive training in specialized query languages. Meanwhile, principles derived from High-Performance Computing (HPC) have emerged in graph analytics platforms to bring supercomputing power to analysts on their desktops.

Graph analytics initiatives no longer require huge up-front investments in tech or training. Organizations of all sizes can use graph analytics to detect anomalies or fraud, create sophisticated forecasts, accelerate research, and create the context for more intelligent agentic AI.

FINDING PATTERNS AND CONNECTIONS IS NOW EVERYBODY'S BUSINESS

Today's businesses and government entities operate in an environment full of interconnected relationships and technology infrastructures. As such, they face a growing list of challenging insight discovery problems that involve uncovering patterns and connections across massive quantities of data. Emerging AI applications—including agentic workflows that must synthesize diverse data points to perform reasoning and take action—are intensifying insight discovery needs. As organizations automate both forecasting and process execution, cybersecurity and fraud detection have become truly mission-critical operations. No cybersecurity team wants to put their company in a position to become the next **Change Healthcare** that brings an entire business ecosystem to a standstill.

Organizations across industries have recognized the need for new discovery and analysis capabilities for cybersecurity, fraud detection, and advanced forecasting. If your organization hasn't yet adopted **graph analytics**, you're missing critical opportunities to advance your business capabilities and protect your operations.

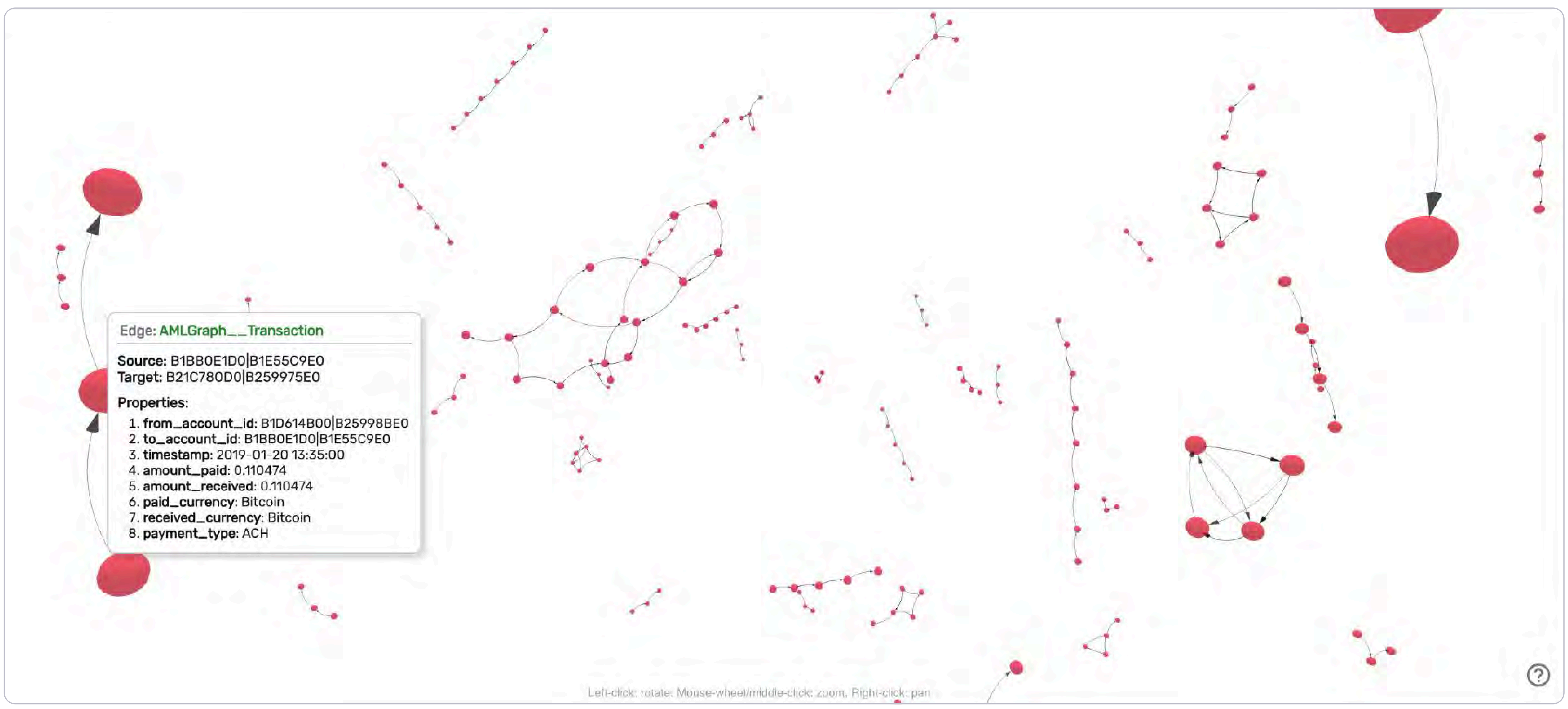
GRAPH ANALYTICS 101

Graphs are an effective way to represent complex systems such as financial networks, communication infrastructures, shipping logistics, criminal networks, social interactions, geographic maps, and molecular structures. A graph consists of nodes (entities within the system) connected by edges (relationships between those entities). Analyzing graphs involves navigating through nodes and edges to identify relationships or view pathways through interconnected data. By modeling data as interconnected nodes and edges, graphs efficiently reveal patterns and relationships that traditional methods may overlook.

The technology to visualize and analyze graphs has evolved significantly over recent decades. Graph analytics originated from cartography models used to navigate cities. Over time, it has expanded broadly to address analytical needs in diverse industries and disciplines.

The most common graph analytics use cases today include cybersecurity, fraud detection, anti-money laundering, counterterrorism, and predictive analytics. In anti-money laundering, for example, criminals attempt to obscure transactions by transferring funds repeatedly across seemingly unrelated accounts. Graph analytics can detect these hidden pathways by finding the connecting data points across transactions. In cybersecurity, analysts use graph analytics to detect behavioral patterns associated with known intrusion methods, such as lateral movement within networks over extended periods.

> According to IBM's **Cost of a Data Breach Report 2024**, the average time to identify and contain data breaches involving stolen credentials was 292 days. The global average cost of a data breach was $4.88 million.



Example of a graph visualization of anti-money laundering connections.

AI BRINGS NEW GRAPH ANALYTICS NEEDS

New use cases are emerging from AI initiatives. These include:

- **GraphRAG** (Graphs + Retrieval Augmented Generation): Graphs enhance Retrieval Augmented Generation (RAG) by using external data sources to enrich the context and improve the quality of AI-generated responses
- **Knowledge Graphs**: Also known as semantic networks, knowledge graphs can support systems from search engines to recommendation models. (IBM has a useful resource for learning more.)

The most widely known graph analytics technology is the graph database, which is structured differently from traditional relational databases. Relational databases store data in tables with rows and columns, which allows for structured data and relationships through joins. They are optimized for structured, transactional data management with atomicity, consistency, isolation, and durability (ACID) properties that keep the database in a valid state even while handling unexpected errors. Graph databases store data as nodes and edges, emphasizing relationships and allowing for efficient traversal of connections.

FIVE EMERGING AI APPLICATIONS USING KNOWLEDGE GRAPHS

1. **Agentic AI**: Knowledge graphs are essential components of critical workflows that you are automating with Agentic AI. They allow your workflows to measure the next steps that AI has inference against your organization's "base truth" to prevent dangerous hallucinations
2. **Cybersecurity**: Go back through years of log data to find potential attack patterns
3. **Precision medicine**: Design treatments that consider diverse biomedical data, including molecular and genetic factors
4. **Drug discovery**: Analyze relationships between chemical structures, genetics, and clinical trial data to find new drugs and new indications–including potential treatments for rare diseases
5. **Supply chain management**: Model complex supply chains to uncover hidden dependencies and potential bottlenecks for contingency planning

SKILLS AND LATENCY: THE TWO BARRIERS TO GRAPH ANALYTICS ADOPTION

Graph analytics effectively addresses a wide variety of complex insight discovery problems. However, two significant barriers have historically limited its widespread adoption. First, the need for deep technical expertise has restricted the use of graph analytics tools to specialist teams with domain expertise to design data schemas as well as the coding skills to write queries. Second, traditional graph databases employ a "scale-out" architecture that limits the type of queries that users can perform and the size of the datasets. They perform adequately for basic graph lookups such as those used in credit card fraud detection. Yet they fall short when handling complex, multi-hop queries for use cases like cybersecurity or anti-money laundering. When queries demand graph traversal with ten or more hops, the traditional graph databases can take weeks (literally!) to produce an answer or not produce an answer at all.

Fortunately, advancements in graph analytics technology have eliminated both of these barriers.

GenAI SIMPLIFIES GRAPH ANALYTICS USERS' WORK

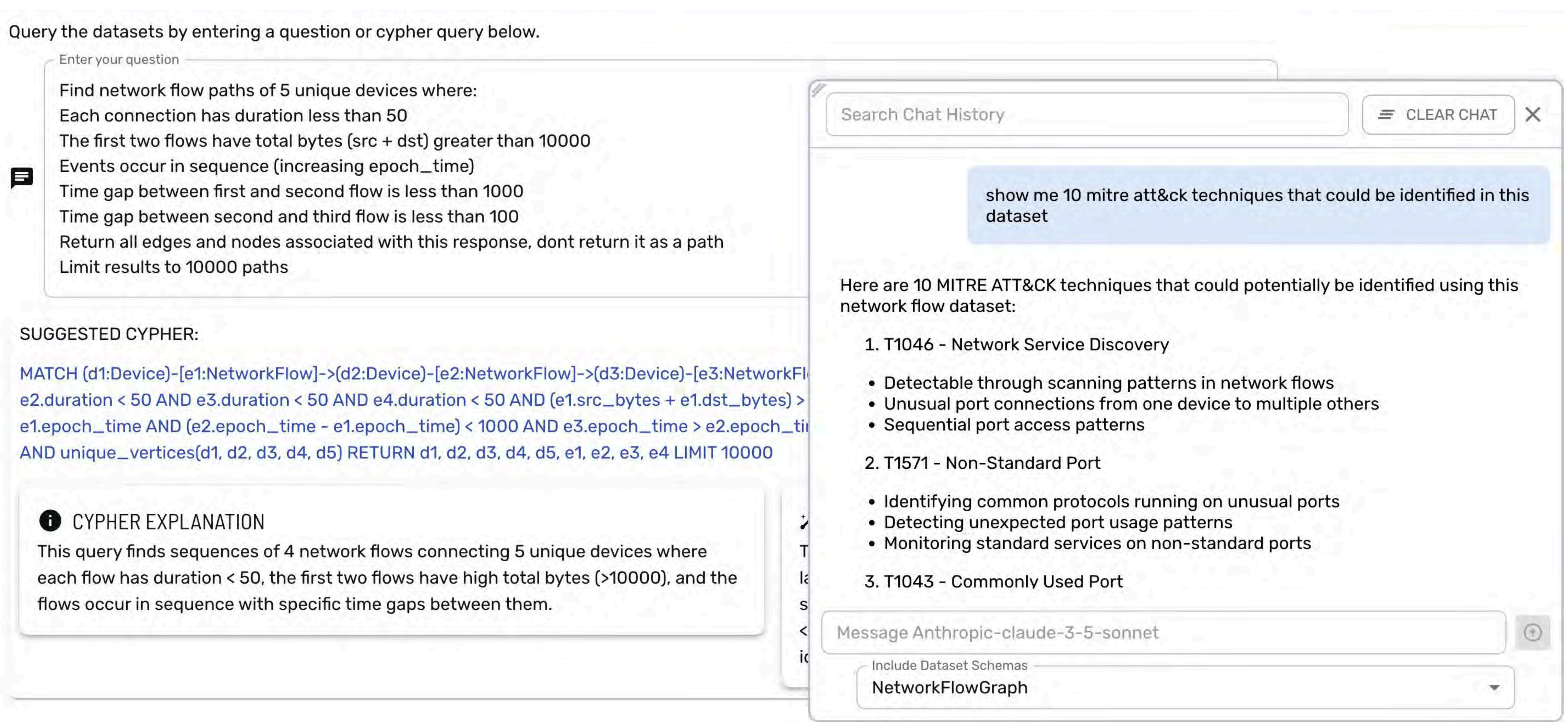
The requirement for specialized expertise is one of the most significant barriers preventing widespread graph analytics adoption. Analysts needed skills in languages like Cypher®, Gremlin, or SPARQL to create data schemas and run queries. Consequently, adoption has remained confined to niche applications within cybersecurity, fraud detection, and national security.

Today, graph analytics technologies are adding large language model interfaces, which allow users to ask questions in everyday language and see results in a graph format. They still need domain expertise to guide their questions, but GenAI can do the heavy lifting behind the scenes.

GENAI IN ACTION: CYBERSECURITY

The MITRE ATT@CK catalog describes numerous methods network intruders use. A patient intruder might spend months or years quietly moving within a network, a technique known as lateral movement, leaving subtle clues along the way. Without writing any code, a security analyst can now leverage GenAI-powered graph analytics to identify instances of specific MITRE ATT@CK patterns. The analyst can simply reference a pattern like T1550.002 in natural language, and the graph analytics platform visualizes all relevant occurrences, significantly enhancing threat detection capabilities.

1. Find examples of T1550.002 patterns using the underlying NTLM or LM hash in my dataset.
2. Make the pattern be a two-hop sequence of lateral movement.
3. The second hop should occur after the first hop.
4. Each hop should carry a significant amount of data.
5. All nodes should be unique.
6. Return all nodes and edges in the pattern, not just a path variable.
7. Order by the dest node



GenAI can convert the analyst's natural language request into Cypher code.

With GenAI integration, virtually every analyst can become a graph analyst. They can ramp up quickly, work productively, and bring in new data sources without having to build even more new skills.

The best graph analytics solutions can also take advantage of your own data and in-house LLM—if you have one—to bring a domain-specific understanding to each analysis while keeping all your data proprietary. They use AI to recommend and generate the graph model that suits your data. For example, a pattern like the MITRE ATT@CK catalog is available with public LLMs, but your private LLM would encode your organization's unique knowledge and intellectual property. This additional context maximizes the accuracy and relevance of insights generated through the platform.

HPC ARCHITECTURES OVERCOME GRAPH ANALYTICS LATENCY ISSUES

With their scale-out architectures that distribute large datasets across multiple nodes, traditional graph databases suffer from severe network latency and performance degradation as datasets grow and queries become more complex. In addition, graph databases are still databases, so they must perform housekeeping tasks to manage ACID guarantees and concurrent users. These tasks further slow down their workloads.

A new generation of graph analytics platforms has adopted the principles of high-performance computing (HPC). Instead, it employs scale-up architectures that put data in-memory on a single node. These platforms dramatically reduce query latency, efficiently supporting datasets with billions of nodes and edges. They employ multithreaded parallelism and compilation optimized for modern processors, sophisticated memory management techniques, and algorithmic innovations inspired by supercomputing principles. Analysts can now access supercomputing power on their desktops, running complex queries in a virtual machine.

The HPC approach makes it practical to deploy graph analytics for the biggest problems: looking through years of financial transactions, network traffic, etc.

CHOOSING THE RIGHT GRAPH ANALYTICS PLATFORM

As your organization evaluates graph analytics solutions, several key considerations should inform your decision-making process:

- **Ease of data integration**: How quickly can you incorporate diverse data sources?
- **Ease of schema creation**: What skills are required to create the schemas that structure your data?
- **Ease of analysis**: Can analysts easily query the platform using natural language, or are complex coding skills required?
- **Scalability**: Has the platform been proven to scale to the most extensive datasets?
- **Performance**: Is the platform designed with supercomputing principles, ensuring performance and reliability as your usage and complexity grow?
- **Competitive advantage**: To what extent can you take advantage of your existing investments in company-specific LLMs?

Platforms that combine GenAI with scale-up architectures offer distinct advantages that result in broader adoption and deeper analytical capabilities for your organization.

THE FUTURE OF GRAPH ANALYTICS STARTS NOW

You might think that graph analytics initiatives require a huge investment to bear fruit; but today, organizations no longer need dedicated teams of highly specialized experts to benefit from these powerful insight discovery tools. Analysts with a wide range of skills can detect anomalies or fraud, create sophisticated forecasts, accelerate research, and create the context for more intelligent agentic AI.

Visit **Rocketgraph** to schedule a demo and see first-hand how GenAI and HPC come together to make widespread graph analytics adoption possible.

ABOUT ROCKETGRAPH

Rocketgraph enables enterprises and government agencies to discover the hardest-to-find insights without hiring a command center full of rocket scientists. Born out of a high-performance computing project at the Department of Defense, our graph analytics platform allows an analyst to use GenAI to do iterative analysis with the largest, most complicated datasets on the planet and get answers hundreds of times faster than traditional graph tools. Rocketgraph builds property graphs that scale to hundreds of billions of edges. Our government and enterprise customers build fine-tuned forecasts, detect sophisticated fraud schemes, monitor nefarious activity on the dark web, keep their networks secure, and answer their most challenging questions with Rocketgraph graph analytics.