neo4j

What's New in Neo4j Graph Data Science

April 2022

What Is Neo4j Graph Data Science?

Neo4j Graph Data Science is a graph analytics and modeling platform. Using graph algorithms and machine learning (ML), data scientists identify patterns and behaviors to improve their models for use across recommendation engines, fraud detection, route optimization, and customer 360 scenarios.

Areas of Investment

We continue to build on the momentum of previous release highlights that included: a fully featured graph workspace, the most graph algorithms on the market, and easy-to-use ML pipelines. Our investments are focused on building the most comprehensive graph data science solution on the market.

We're investing in five key areas:

1. **Easy to use:** Take the fastest path to production.
2. **Graph built for data scientists:** Analyze graphs through a library of graph algorithms, ML pipelines, and data science methods.
3. **Enterprise ready:** trusted, scalable, and robust: Securely handle hundreds of billions of nodes and relationships.
4. **Data pipeline ecosystem:** Integrate Graph Data Science with the existing tools across your technology stack and data pipeline.
5. **Graph data science as a service:** Get a fully managed SaaS solution.

What's new?

Easy to use

Improvement of experiences to make them simple, fast, and intuitive include:

1. **Improvements to similarity algorithms:** K-Nearest Neighbors and Node Similarity support the full catalog of similarity metrics to simplify property and topological similarity measurements.
2. **Unified ML pipeline:** ML pipelines (link prediction, node classification) have a single syntax, streamlined workflow, and can be saved for reuse or published and shared with your teams.
3. **Simplified Cypher projections:** Use a single command to experiment with the right schema for your use case while retaining the integrity of your database.
4. **Hints and warning messages:** A new procedure, GDS.alpha.user log, to automatically log warnings and hints about your code.

Graph built for data scientists

Features that empower data scientists to analyze graph are:

- **Native Python client:** Easily call Graph Data Science procedures from Python without setting up and configuring a driver or transaction functions.

- **Random Forest:** Node classification and link prediction now support the classification algorithm, Random Forest, in addition to logistic regression.

- **Algorithms graduating to product (generally available):**

  - Delta stepping shortest path algorithm calculates the shortest (weighted) path from a node to all other nodes in the graph. Delta stepping is a parallel implementation, ensuring speed and scalability.

  - The K-Nearest Neighbors algorithm computes a distance value for all node pairs in the graph and creates new relationships between each node and its K-Nearest Neighbors. The distance is calculated based on node properties like embeddings,with new features, like similarity thresholds, support for multiple node properties, and improved accuracy.

  - The Breadth First Search algorithm is a graph traversal algorithm that, given a start node, visits nodes in order of increasing distance.

  - The Depth First Search algorithm is a graph traversal that starts at a given node and explores as far as possible along each branch before backtracking.

Enterprise ready: trusted, scalable, and robust

Features that support enterprise volumes of data, use cases, and complexity include:

- **Graph backup and restore:** Back up your graphs to disc and restore them following a database restart.

- **Cluster compatibility:** Automatically sync your transactional clusters with a dedicated Graph Data Science instance with no ETL.

- **Graph compression and loading improvements:** Decreased load times and memory requirements – so you can analyze bigger datasets faster.

Data pipeline ecosystem

New and improved connectors, extensions, and integrations across the data pipeline ecosystem include:

- **Neo4j Spark Connector:** The Neo4j Spark Connector is fully supported and can be used with Amazon Redshift, Databricks, Google BigQuery, Microsoft Azure Synapse Analytics, and Snowflake. Connect your graph with your enterprise data warehouse or to all the ML capabilities in MLflow.

- **Neo4j BI Connector:** Give direct access to BI tools like Microsoft Power BI, Tableau, Looker, TIBCO Spotfire, Microstrategy, and more through the Neo4j BI Connector.

- **Dataiku:** Use the Dataiku plugin to include Neo4j in your ML workflows.

- **KNIME:** Drag and drop Neo4j nodes into a KNIME pipeline with the KNIME extension.

Graph data science as a service (AuraDS)

AuraDS is generally available: Use our fully hosted graph data science service to import, analyze, and visualize graph data through a SaaS license that includes all aspects of the infrastructure stack: storage (including backup storage), IO rate, and data transfer. AuraDS includes Graph Data Science Enterprise Edition for data science graph algorithms and the modeling workspace and Bloom for visual data exploration, all supported by a graph database.

Features in our fully managed cloud offering include:

- **Google Cloud Platform support:** AuraDS is available first on Google Cloud Platform. For guidance and reference architectures on how to use AuraDS with Vertex AI, see our latest blog, Use graphs for smarter AI with Neo4j and Google Cloud Vertex AI

- **Neo4j Data Importer:** Use a drag and drop UI to model and import data into a graph.

- **Scale up and down on demand:** Manage your access to high compute hardware as your needs change.

- **One-click backup:** Take a snapshot of your instance, models, and in-memory graphs with one click.

- **Automated operations:** Workloads are monitored, patched, and backed up behind the scenes without any user action.

- **MLOps support:** Persist, publish, and restore models without interruptions from restarts.

- **Predictable and simple cost management:** Manage your costs with pay-as-you-go pricing and the option of pausing unused instances to pay less.