

A brief introduction to Neo4j

Neo4j (<https://neo4j.com/>) is an interactive, native graph database, purpose-built to leverage data relationships. The objects studied in each analysis are the typical components of a graph, such as nodes, edges, paths between the nodes, node centrality and what kind of relation links a node to another one. In Neo4j, everything is stored in the form of an edge, node, or attribute. Each node and edge can have any number of attributes. Both nodes and edges can be labelled. Labels can be used to narrow searches.

Neo4j allows us to run queries written using the **Cypher** query language, over a data model expressed as graph and working with the results obtained in both graph visualization and tabular forms. If the Cypher query returns graph entities (nodes, relationships, paths, etc.), then **Neo4j Browser** will render a graph visualization able to be explored. However, often the answer to a question is returned in tabular data, for instance, the result of running an aggregation. It is possible also to return rows of values, in that case Neo4j Browser will render a table of rows.

Introductory tutorials are available at <https://neo4j.com/docs/getting-started/>

Introduction to Cypher is available at <https://neo4j.com/docs/getting-started/cypher-intro/>

Neo4j Sandbox

The **Neo4j Sandbox** enables you to get started with Neo4j, with built-in guides and sample datasets for popular use cases. The sandbox comes pre-loaded with sample data and a step-by-step guide with queries and explanations. You can start a free sandbox from <https://neo4j.com/sandbox/>. Besides the Movies sandbox that we used during class, the available sandboxes with pre-loaded data (and classified as “For Developers” and “For Data Scientists”, but do not worry about this classification) are:

- **ICIJ FinCEN Files Investigation:** Explore large volumes of Suspicious Activity Report (SAR) filings between entities around the globe using the [FinCEN Files](#) investigation.
- **Network and IT Management:** Dependency and root cause analysis plus more for network and IT management.
- **Crime Investigation:** Explore connections in crime data using the POLE (Person, Object, Location, Event) model in a public dataset from Manchester, U.K.
- **Graph Data Science:** Leverage [Neo4j Graph Data Science](#) library to explore graph algorithms for analytics and feature engineering; the dataset is based on the Game of Thrones fantasy saga.
- **Women's World Cup 2019:** Explore the data behind the Women’s World Cup with the World Cup Graph.
- **Recommendations:** Generate personalized real-time recommendations using a dataset of movie reviews. This sandbox also uses [GraphQL](#), another query language.
- **OpenStreetMap:** Identify points of interest and routing with Neo4j using the [OpenStreetMap](#) dataset of Central Park in New York City. This sandbox also uses [GraphQL](#), another query language.
- **Fraud Detection:** Identify fraud detection with the Paysim financial dataset, Neo4j Graph Data Science, and Neo4j Bloom.
- **Contact Tracing:** Explore contact tracing using a synthetic dataset of places, persons, and visits.
- **Paradise Papers by ICIJ:** The Paradise Papers dataset and guide from the International Consortium of Investigative Journalists (ICIJ).

- **ICIJ Offshoreleaks:** The [Offshore leaks](#) dataset and guide from the International Consortium of Investigative Journalists (ICIJ).
- **Russian Twitter Trolls:** Explore data released by NBC News from their investigation into Russian Twitter Trolls around the 2016 US election.
- **Twitch:** Explore data related to [Twitch](#) social network, an online platform that allows users to share their content via live stream.
- **Stack Overflow:** Explore data related to [Stack Overflow](#) questions, answers, tags, and comments and the relationships between them.
- **Entity Resolution:** Entity Resolution is the process of disambiguating data to determine if multiple digital records represent the same real-world entity such as a person, organization, place, or other type of object. This use case is about entity resolution for an online movie streaming platform.
- **Cybersecurity:** This sandbox is based on the data and themes from the [BloodHound](#) project, which is a tool for auditing an Active Directory environment. Active Directory helps IT teams monitor various network resources and users and allows to grant and revoke different user permissions. Using Neo4J and [Neo4j Graph Data Science](#) library, you can analyze all possible attack paths based on access.
- **Healthcare Analytics:** Load and analyze FDA Adverse Event Reporting System data with Neo4j. The FDA Adverse Event Reporting System is an information database designed to support the U.S. Food and Drug Administration's post marketing safety surveillance program for all approved drug and therapeutic biologic products.
- **Contact Tracing:** Explore contact tracing using a synthetic dataset of places, persons, and visits.

By default, each sandbox you create is available for **3 days**, but you have the option to extend it for more 7 days (making it a total of 10 days), using a link that Neo4j will send you by email 24 hours before it expires.