Introduction to Graph Databases

Alejandro A. Vaisman

Instituto Tecnológico de Buenos Aires
avaisman@itba.edu.ar

1 Goals

To introduce the students the main concepts related to graph databases (GDB), with emphasis in Neo4j.

2 Course Description

The first (on-site, face-to-face) part of the course will introduce the context for GDB, and how they situate within the NoSQL paradigm. The main concepts, tools, and techniques for GDB will be studied, with emphasis in the property graph data model and Neo4j (and its companion query language, Cypher). Graph data science will also be studied. Finally, the RDF data model, and RDF triple stores will be covered, as an alternative to the property graph data model.

3 Learning Outcome

At the end of the course, the students will be able to model and query a GDB, and to evaluate the convenience or not of using such database instead of (typically) a relational database, for a given problem.

4 Course Content

1. Day 1. (31.03.2025). Introduction to Big Data and the NoSQL paradigm. Fundamentals of graph databases. The property graph data model. Property graph databases vs. Relational databases.
2. Day 2. (01.04.2025). Neo4j data model. The Cypher query language. Basic and advanced queries. Advanced queries in Cypher.
3. Day 3. (03.04.2025). Graph Data Science in Neo4j. Content and Collaborative Recommendation Systems. Examples: Twitter, Twitch, OSM, etc.
4. Day 4. (04.04.2025). Knowledge Graphs. RDF graph data stores. Property graphs vs RDF graph stores. The Semantic Web. The SPARQL Query Language. Comparing Property Graphs vs RDF Graphs. Graph semantics.
5. Offline Project. There will be a course project to be done offline.
   Deadline: TBD.

5 Assessment Method

The final course grade will be the weighted average of the grades of the two parts of the course project. To pass the course students must attend a minimum of three lectures.

6 Prerequisites

Good knowledge of relational databases and SQL.

7 Bibliography

1. R. Angles. A Comparison of Current Graph Database Models. In *Proceedings of ICDE Workshops*, pages 171–177, Arlington, VA, USA, 2012.
2. Renzo Angles and Claudio Gutierrez. Survey of graph database models. *ACM Comput. Surv.*, 40(1):1:1–1:39, 2008.
3. NoSQL Databases. http://nosql-database.org/.
4. O. Hartig. Reconciliation of RDF* and property graphs. *CoRR*, abs/1409.3288, 2014.
5. P. Hitzler, M Krötzsch, S Rudolph. *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC, 2009
6. Grzegorz Malewicz, Matthew H. Austern, Aart J.C Bik, James C. Dehnert, Ilan Horn, Naty Leiser, and Grzegorz Czajkowski. Pregel: a system for large-scale graph processing. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pages 135–146. ACM, 2010.
7. Neo4j. http://neo4j.org/.
8. Ian Robinson, Jim Webber, and Emil Eifrem. *Graph Databases*. O'Reilly Media, Inc., 2013.
9. E. Scifo. *Graph Data Science with Neo4j*. Packt Publishing, Birmingham - Mumbai, 2023.
10. A. Vaisman and E. Zimányi. *Data Warehouse Systems: Design and Implementation, 2nd. Edition (Chapter 13)*. Springer, 2022.