CAPCOU

Knowledge graphs

Transforming messy data into actionable insights With huge volumes of information being generated and processed every second, it is paramount in today's fast-evolving financial services landscape that institutions can access clean, governed data to ensure they are making educated and prudent decisions. To achieve this goal, many financial institutions are turning to knowledge graphs.

As outlined in <u>our previous paper</u>, a knowledge graph is a semantic approach to formatting and storing data across multiple data silos within different technology stacks. Instead of organizing data in traditional rows and columns, a knowledge graph breaks it down into nodes and the relationships between them, called 'triples'.

However, when transitioning to knowledge graphs from traditional legacy databases, understanding where and how to start can be a challenge. In this article we provide an overview of the principal considerations to mobilize your knowledge graph initiative – understanding the key concepts, establishing the prerequisites, and moving into implementation and scaling.

0

Knowledge graphs: Transforming Messy Data Into Actionable Insights /2

Understanding the key concepts

There are three main components to creating a knowledge graph:

- Ontology this is the blueprint that defines how all your data concepts relate to each other
- Instance data these are the triples created in accordance with your ontology
- Querying this involves interrogating the data and generating reports from your knowledge graph.

Establishing the prerequisites

In order to mobilize and begin to drive your knowledge graph initiative, these essential prerequisites should be in place:

- Prepare your data environment
- Select your preferred tooling and technology options
- Assemble an appropriate team.

Preparing your data

One of the major benefits of a knowledge graph is the ability to connect many disparate datasets. To achieve this, you do not need to change and move all your data out of their current storage and toolsets. Rather there are two main options available: virtualisation and Extract Transform Load (ETL).

 Virtualisation – many knowledge graph tools enable you to virtualize data from different sources on top of your existing knowledge graph. This approach allows you to keep data in memory rather than converting it from traditional relational databases to RDF triples and storing them on disk. Virtualization facilitates quick experimentation with various data sources and rapid querying and results, bypassing the need for labour-intensive pipelines or obtaining permissions to store data in a new location.

Extract Transform Load - ETL refers to the process of extracting data from traditional data sources, converting it into an RDF triple format, and storing it in a triple store. This approach allows you to retain your traditional data storage systems while benefiting from having your data in a knowledge graph.
ETL is the most common methodology used, with numerous tools and techniques available for implementation.

Virtualization and ETL have their respective advantages. In a knowledge graph, you can mix and match these methodologies for different data sources, allowing you to leverage the benefits of both approaches.

Selecting your tools

While knowledge graphs are still gaining traction within the financial services sector, they have already been widely adopted in other industries. A diverse array of tools are accordingly available to meet individual needs and budgets. Rather than reviewing each individual tool, this paper highlights the advantages of adopting an RDF triplestore for your knowledge graph infrastructure. Because RDF is a W3C-standard data model and SPARQL its universal query language, you gain seamless interoperability across platforms. This approach eliminates vendor lock-in and gives you the flexibility to choose-and switch between-the most appropriate tools at every stage of your knowledge graph journey.

Some tools prioritize the speed of data loading, some focus on front end visualizations, while others emphasize interoperability with existing front end tools and data sources. Tool selection should be guided by your specific requirements: if you require a high-performance backend that delivers the fastest query response times, a specialized triplestore engine is most appropriate; whereas if you need a comprehensive end-toend platform with built-in data visualization and exploration capabilities, a more fully featured solution will better serve your needs. Additionally, there are completely free tools available that allow you to explore whether a knowledge graph solution is the right fit for your organization.

To explore the latest Knowledge Graph tools and identify the solution best suited to your needs, please contact Capco. Our experts will guide you in selecting the most effective tool for your objectives.



Assembling your team

Building knowledge graphs does not require a small army. Below we provide an overview of the different roles involved (also see Figure 1).

- **Ontologist** designs the mapping of the data and how different data sets are related
- **Graph Engineer** manages the pipelines and transformations of your data from your current data sources to a knowledge graph
- Graph Analyst queries your knowledge graph and generates reports
- **Graph Data Scientist** constructs data transformation pipelines and queries, but has little experience in modelling data
- Graph Architect able to create data transformation pipelines, and has experience in data schemas and modelling, but little experience querying or creating reports
- Solutions Architect often involved in modelling the data and creating reports, allowing for iterative ontology design, though not involved in the more technical data transformation

• Knowledge Graph Engineer – able to create data transformation pipelines regardless of your technology stack, is experienced in modelling data, then querying that data and creating reports.

Deciding on the right team mix depends on your organization's needs. Some may require only a Graph Data Scientist and an Ontologist, while others may need a broader range of skills.

Striking the right balance is crucial, however. Having too many Ontologists can lead to indecision around modelling, for example, whereas relying solely on Graph Data Scientists might result in a technology-driven rather than an ontologically-driven model.

Capco can provide further information on how to structure a knowledge graph project to meet your specific requirements.



Figure 1

Implementation approach

Given the iterative and evolving nature of a typical knowledge graph project – with efforts to solve one problem often leading to two or three further issues being identified and resolved – how you choose to structure your team and lock down the processes required to scale up are crucial considerations. It is usually most effective to create a central Ontology design team, with silos directed at each data source and required output. This approach ensures that – regardless of your starting point, whether addressing one problem or ten – your ontology remains consistent, your data remains governed and interconnected, and your team does not become overworked.



Figure 2

Initial Deployment

The duration of a knowledge graph project will depend on the specific requirements and the size of the team deployed. These projects are highly iterative and can scale quickly, but at the outset require significant effort to establish.

As outlined in our previous paper, you need to design an ontology before transforming and connecting your data and querying it. This means there is a substantial initial inertia that must be overcome. Based on our experience, a proof of concept (POC) project involving a few data sources typically requires a team of three and approximately six months to complete.

After that, each additional report, requirement, or data source can take anywhere from one month to six months, depending on the complexity, format, and location of the data.

Further Integration and Scaling

Once the initial problem is solved or the initial request is met, the parameters of the project often expand.

This growth may involve increasing the number of data sources feeding your graph or the number of stakeholders relying on the data. These are ideal challenges for a knowledge graph to address, though they often require an increase in team size.

However, organizations often encounter the following pitfalls:

- **Duplicated efforts** multiple models describing the same data within your organization
- Unconnected ontologies inability to connect multiple ontologies effectively if they are not built to the correct W3C standards¹

 Confusing queries – departments using different terminology for the same concepts without proper communication can lead to misinterpretation within the knowledge graph.

To avoid these pitfalls, organizations should adhere to a centralized ontology team model. This team manages all new knowledge graph projects and their integration with the established knowledge graph.

Smaller teams can create proof-of-concept knowledge graphs with different departments, coordinating with domain experts in those areas. These POCs are then integrated back into the main knowledge graph. This approach allows multiple departments to create interconnected, insightful, and governed knowledge graphs that enhance overall utility and insight.

At Capco, we have worked globally on large enterprise knowledge graphs with multiple leading financial institutions. Our highly skilled team of experts can guide you through your knowledge graph journey including creating a team, choosing a tool, and outlining a project timeline. Please contact us for more details.

^{1. &}lt;u>https://www.w3.org/</u>

Authors

Caleb Severn Rollo Burgess

Contact

Rollo Burgess Partner rollo.burgess@capco.com

About Capco

Capco, a Wipro company, is a global management and technology consultancy specializing in driving transformation in the energy and financial services industries. Capco operates at the intersection of business and technology by combining innovative thinking with unrivalled industry knowledge to fast-track digital initiatives for banking and payments, capital markets, wealth and asset management, insurance, and the energy sector. Capco's cutting-edge ingenuity is brought to life through its award-winning Be Yourself At Work culture and diverse talent.

To learn more, visit www.capco.com or follow us on LinkedIn, Instagram, Facebook, and YouTube.

Worldwide Offices

APAC

Bengaluru - Electronic City Bengaluru - Sarjapur Road Bangkok Chennai Gurgaon Hong Kong Hyderabad Kuala Lumpur Mumbai Pune Singapore

MIDDLE EAST Dubai

зт

EUROPE

Berlin Bratislava Brussels Dusseldorf Edinburgh Frankfurt Geneva Glasgow London Milan Paris Vienna Warsaw Zurich

NORTH AMERICA

Charlotte Chicago Dallas Houston New York Orlando Toronto

SOUTH AMERICA São Paulo



© 2025 The Capital Markets Company (UK) Limited. All rights reserved.

