



Canada Revenue
Agency

Agence du revenu du
Canada

REQUEST FOR INFORMATION (RFI)

NO. 1000405070

**GRAPH DATABASE AND VISUALIZATION SOLUTION
FOR THE CANADA REVENUE AGENCY**

Closing date and time : September 8th, 2022 - 2:00 PM (EDT)

DISCLAIMER

Responding to this Request for Information (RFI) is not a prerequisite to receiving or being eligible to bid on any Request for Proposal (RFP) for this requirement. Any RFP will be advertised on the Government Electronic Tendering Service (GETS) commonly referred to as Buy and Sell (<https://buyandsell.gc.ca/>).

This RFI is not to be construed as a solicitation for tenders or proposals. No contract or other form of commitment will be entered into based on responses to this RFI. This RFI is not considered as authorization by the Canada Revenue Agency (CRA) to undertake any work that would result in costs to CRA.

Nothing in this RFI shall be construed as a commitment from CRA to issue an RFP for this program. CRA may use non-proprietary information provided in its review and/or in the preparation of any formal RFP. All responses will be held by CRA on a confidential basis (subject to applicable federal legislation) and remain the property of CRA once they have been received.

CRA may reproduce or photocopy or transcribe the response and any non-proprietary supporting documentation for the purpose of its review and/or inclusion in any resulting RFP document. Vendors responding to this RFI are advised to clearly identify which (if any) portions of their responses are proprietary and may be invited to a meeting to further clarify their responses to questions provided in Appendix A and B herein.

CRA shall not be bound by anything stated herein. CRA reserves the right to change, at any time, any or all parts of the requirements, as it deems necessary. CRA also reserves the right to revise its procurement approach, as it considers appropriate, either based upon information submitted in response to this RFI or for any other reason it deems appropriate.

Responses to this RFI will not be used to pre-qualify or otherwise restrict participation in any future procurement process (e.g. an RFP). Responses will not be formally evaluated.

CRA will not reimburse any expenditure incurred in preparing responses and participating in the presentation sessions related to this RFI.

Interactive Demonstration Sessions

CRA may at its sole discretion entertain presentations/demonstrations with interested respondents who have clearly addressed the Solution Requirements in their response to CRA to provide them with the opportunity for a follow-up to their written response to present their capabilities in relation to this RFI.

Respondents may be contacted within 20 business days of the RFI closing date to schedule the presentations/demonstrations. Specific questions or areas of interest to be covered during the session may also be provided and will be based on responses received.

Presentations/demonstrations will be virtual utilizing MS Team.

The time frame allocated for each session will be a maximum of 2 hours.

Respondents must be familiar with the services capabilities to respond to questions at the presentation/demonstration session.

Responses and Enquiries

Responses for questions must be submitted complete and in writing in the order shown. All requests for information in all sections of this document must be answered as concisely as possible while providing all information necessary to understand the proposed solution. Any deviation from the question or requirements that cannot be satisfied by the vendor, must be clearly identified.

Any information of a confidential or proprietary nature contained in a Vendor's response should be clearly marked 'PROPRIETARY' or 'CONFIDENTIAL' by item or at the top of each page.

The Vendor must provide a contact name, email address and telephone number when submitting their response.

Respondents are requested to submit responses by 2:00PM Eastern Daylight Time (EDT), September 8th, 2022. Responses received after this date/time will not be reviewed.

Electronic submissions are mandatory and should be submitted as one complete package.

Vendors are requested to submit responses to this RFI using the following e-mail:

Hamza.Khanqui@cra-arc.gc.ca

All enquiries must be submitted via email to the attention of Hamza Khanqui at

Hamza.Khanqui@cra-arc.gc.ca

CONTEXT

Introduction

The Canada Revenue Agency's Compliance Management Directorate, in the Information Technology Branch is the Agency's technical lead to provide an enterprise solution for the Small, Medium and Enterprise (SME) division, for data analytical tools and techniques that can manage an ever-increasing amount of data, find relationships within the data, and derive value from additional data sources. One of the core components of the project is to have a solution that makes use of graph database technology to find relationships between entities, store and retrieve attributes pertaining to these entities and relationships, as well as the visualization of this data.

The purpose of this Request for Information (RFI) is to gather information on industry capabilities and availability of a Graph Database and Visualization solution that can satisfy our business requirements.

The key objectives of the RFI include:

1. Determine vendor capabilities in providing a cloud-based graph database solution that integrates with or includes advanced visualisation and data science algorithms.
2. Help CRA understand industry standards, best practices, and/or recommendations in terms of using a graph database and graph database visualization tools in a public service environment.
3. Get a better understanding of the current and future trends of Graph Database solutions, integrated with Machine Learning and Advanced Visualisation.
4. Provide an opportunity for industry to demonstrate and discuss its software functionalities, capabilities, and constraints.
5. Solicit feedback on options for integrating the solution with different platforms and systems.
6. Solicit feedback on the licensing model, schedule, level of effort, hardware requirements and technical architecture.
7. Vendors who establish via their response to the RFI how their products(s) meet the solution requirements may be invited to provide an interactive demonstration and discuss in detail how their solutions meet the listed requirements.

Background Information

There is a need for a graph database to hold the structure of Economic Entity's (EE), which are groups of corporations, individuals and trust taxpayers and partnerships controlled by the same person(s). Besides the graphical representation of the nodes (legal entities) and the relationships, it would also include properties for each legal entity type, that allow EEs to be ranked by level of risk and facilitate screening. Since risk assessing taxpayers, as related groups, is more efficient than risk assessing taxpayers one by one, the ultimate goal is to include EE information in the SME risk assessment process to ensure that the highest risk SME taxpayers and EEs are being audited.

Requirements

Graph Database:

1. Handle in excess of >100 million nodes and >300 million edges total
2. Handle a single graph that contains 400,000 nodes and 800,000 edges
3. Support and/or compute embedding, including node, edge, sub-graph embedding
4. Pattern Matching for graph query(for example):
 - a. Collections of subgraphs that fit the conditions
 - b. Return nodes that fit the selection criteria
 - c. Return relationships/edges that fit the selection criteria
5. Export data
6. Able to hold/store in-depth properties/details on each node and edge
7. Ability to query shortest path or limit arms length
8. Cloud hosted capability
9. Ability to scale up new data sources with minimal effort
10. Audit trail capability
11. Role-based access to perform different functions
12. Application programming interface (API) for importing, generating, editing and querying graph.
13. Support parallel processing and load balancing with multiple servers/nodes

Graph Database Visualization (ideally part of the graph database product offering):

1. Cloud hosted capability
2. Support Network Analysis (for example):
 - a. Page rank
 - b. Clustering ratio
 - c. Centrality measures
 - d. Information propagation measures
3. Visual representation that can handle a graph containing 400,000 nodes (panning/collapsing etc.)
4. Able to label nodes and edges
5. Able to view stored details of a node/edge
6. Manipulable visual representation of subgraph that fits predefined conditions
7. Visualizes data in the graph and allows users to navigate and easily query the data without any formal query language or programming knowledge required.
8. Community detection (e.g. Louvain algorithm)
9. Role-based access to perform different functions
10. Ability to trace nodes along with their relationship(s) (for example):
 - a. Filtering the secondary relationship(s) initially to avoid overcrowding display
 - b. Allow to expand and to trace neighborhood nodes by their relationship(s)

Potential Use Cases

Describe how your solution can be used to implement the following use cases:

- Discovering how entities (Corporations/Shareholders/Individuals/Partnerships) are related to each other and group them accordingly
- Find key characteristics of the entities and the relationships between them in order to group/sub-group them and determine their risk

Constraints to be considered for the proposed solution

Official Languages - Must meet the Government of Canada standard under the [Official Languages Act](#), specifically user interfaces, functionality and documentation in English and French.

Accessibility – Must meet the Government of Canada standard under the [Accessible Canada Act](#).

To accomplish this, the CRA has adopted the [EN 301 549 V2.1.2 \(2018-08\) Harmonised European Standard](#) for internally facing Information and Communication Technology (ICT) products and services and the [Web Content Accessibility Guidelines \(WCAG\) 2.0 Level AA](#) for externally facing ICT products and services. Additionally, the CRA's goal is for products and services to be more accessible to and more usable by the broadest range of government officials and Canadians who use them, including those with disabilities. As such, the CRA is planning to adopt the [EN 301 549 V2.1.2 \(2018-08\) Harmonised European Standard](#), which includes the latest version of the Web Content Accessibility Guidelines (WCAG) 2.1 AA (2018), for externally facing ICT products and services as well by March 2021 (final date subject to change).

Privacy - Must meet the Government of Canada standard under the [Privacy Act](#).

User access – Must support limiting user access (user profiles) and configuration.

RFI Questions:

Responders are required to answer RFI questions in Appendix A and B.

Appendix A – General Questions

The following questions are representative of the type of information the CRA is seeking as it considers how to structure any RFP that might follow this RFI process.

Vendors must note that this list of questions is not exhaustive; vendors are invited to provide any additional information that might prove useful and/or beneficial to the CRA in preparing any subsequent RFP.

Vendors that provide written feedback may be invited to a one-on-one consultation session with CRA representatives. This session would allow vendors to provide additional feedback pertaining to this RFI as well as to explain comments made in their written submission.

A.1 – General Information	
A.1.1	Describe how your solution meets the Government of Canada standard under the Official Languages Act, the Accessible Canada Act, and the Privacy Act. <ul style="list-style-type: none"> i. Describe in what capacity your solution provides user interface functionality and documentation in English and French. ii. Describe how your solution supports accessibility, especially persons with disabilities. iii. Describe how your solution meets the Government of Canada standard under the Privacy Act
A.1.2	Provide two examples where your solution was implemented. Include implementation time, common success factors, and obstacles in standardizing this solution.
A.1.3	Any use case of your solution for application in government (preferably in the Canadian Federal Government)?
A.2 – Licensing	
A.2.1	What are the available licensing models (e.g. site licences, concurrent users, data volume, etc.), including maintenance and support models?
A.3 – Training, Documentation, and Support	
A.3.1	What type of technical documentations do you provide for your application (e.g. white paper, installation and configuration instruction, release notes, user guides, etc.)?
A.3.2	Describe the different types of training you provide for this type of solution.
A.3.3	Do you offer standard and customized training?
A.4 – Deployment and Technical Support	
A.4.1	Describe your release schedule, including major and minor enhancements of your software. How often do you release software updates? How does the upgrade impact the implementation of existing services?

A.4.2	Does your solution include 3rd party or open source software? If so, please list them.
A.4.3	Describe the activities and the type/level of expertise required to maintain the solution on an on-going basis.
A.4.4	Describe your maintenance and support offerings (e.g. pre-deployment, post-deployment, consulting, after-hours support, 7/24 on-call support, etc.)
A.5 – Functional	
A.5.1	Describe what functional components are included in the base solution and what additional components are available.
A.5.2	Is there any customization required before the system is operational. Can this be done by a customer or does it require vendor setup?
A.5.3	Describe any special supporting IT infrastructure, capacity, or environmental dependencies required by your solution.
A.5.4	Presuming the necessary IT infrastructure was already in place, how long would it take to install your solution, configure it, and make it ready for production?

Appendix B – Technical Questions

The following questions are representative of the type of information the CRA is seeking as it considers how to structure any RFP that might follow this RFI process.

Vendors must note that this list of questions is not exhaustive; vendors are invited to provide any additional information that might prove useful and/or beneficial to the CRA in preparing any subsequent RFP.

Vendors that provide written feedback may be invited to a one-on-one consultation session with CRA representatives. This session would allow vendors to provide additional feedback pertaining to this RFI as well as to explain comments made in their written submission.

B.1 – General Technical Questions	
B.1.1	Does your solution support a cloud and/or hybrid implementation? How does it integrate with on premise legacy systems?
B.1.2	Describe the architecture of your solution including components that need to be deployed for user interfaces, clustering, high availability, and scalability features.
B.1.3	Describe the footprint/impact that your solution would have in terms of CPU usage, server components, or remote components required. Are there any options that could minimize resource consumption?
B.1.4	Which Cloud Service Provider(s) does your solution support?
B.2 – Graph Database Questions	
B.2.1	Is your solution a native graph storage? What is the data model (RDF, property etc)?
B.2.2	Does your solution support Knowledge Graphs?
B.2.3	Describe which graph query languages your solution supports. Is there any plan to adopt the upcoming GQL Standard (ISO.IEC 39075)? Graph Query Language GQL - What is a GQL Standard? (gqlstandards.org)
B.2.4	Describe your solution's graph algorithm library. Are they open sources? Is the code accessible for software audit, including security vulnerabilities audit?
B.2.5	Which programming language are supported (e.g. Java, python, C++) in regards to your graph algorithm library?
B.2.6	What is the frequency of new graph algorithm library releases. What are the procedures to update Graph Algorithm library?
B.2.7	Does your solution come with a visual interface?
B.2.8	Does your solution support connection to other Relational Database Management System (RDBMS) for importing graph?

B.2.9	Does your solution support point-in-time?
B.2.10	Which character encoding does your solution support?
B.3 – Graph Database Visualization Questions	
B.3.1	Does your solution have the ability to call/invoke external APIs/URLs?
B.3.2	Does your solution have the ability to make use of ArcGIS (or similar functionality)?
B.3.3	Describe your solution’s network analysis options.
B.3.4	Describe your solution’s output options.
B.3.5	Can data be modified using your solution? (remove/add a node/edge, change a label etc.)
B.3.6	Which graph databases does your visualization solution support?
B.3.7	Does your solution support graph layout such as Hierarchical Graph Layout. What other graph layouts does your visualization solution also support?
B.4 – Security, Audit, Logging	
B.4.1	Describe your solution’s security features in detail.
B.4.2	Describe your solution’s authentication and authorization options.
B.4.3	Explain how your solution can integrate with existing authentication and authorization systems.
B.4.4	Describe how security patches and fixes are implemented on your solutions.
B.4.5	Provide all available Security certifications and assessments for the proposed solution, if any.