

Scalable, Secure Associative Database

Navy SBIR 2013.2 - Topic N132-131

ONR - Ms. Lore Anne Ponirakis - loeanne.ponirakis@navy.mil

Opens: May 24, 2013 - Closes: June 26, 2013

N132-131 TITLE: Scalable, Secure Associative Database

TECHNOLOGY AREAS: Information Systems, Sensors, Human Systems

ACQUISITION PROGRAM: FNT-FY12-02 Autonomous Persistent Tactical Surveillance, DCGS-N ACAT IAM

RESTRICTION ON PERFORMANCE BY FOREIGN CITIZENS (i.e., those holding non-U.S. Passports): This topic is "ITAR Restricted". The information and materials provided pursuant to or resulting from this topic are restricted under the International Traffic in Arms Regulations (ITAR), 22 CFR Parts 120 - 130, which control the export of defense-related material and services, including the export of sensitive technical data. Foreign Citizens may perform work under an award resulting from this topic only if they hold the "Permanent Resident Card", or are designated as "Protected Individuals" as defined by 8 U.S.C. 1324b(a)(3). If a proposal for this topic contains participation by a foreign citizen who is not in one of the above two categories, the proposal will be rejected.

OBJECTIVE: Develop and demonstrate an open-source, non-proprietary, scalable, and secure associative database for enhanced discovery of relationships, clues, and evidential insights that are hard to find in a range of missions that address asymmetric threats and unlawful activities.

DESCRIPTION: Analysts supporting naval missions have to develop actionable intelligence from an extensive amount of data that require a multi-disciplinary approach for automated processing of structured and unstructured data types from an expanding array of sensors and information sources. Such automated technologies will significantly reduce the time to develop appropriate measured response when involved with rapidly evolving events or missions such as counter-terrorism activities, counter-narcotics operations, or reconnaissance of agile enemy forces. Associative database techniques provide analysts with high-performance discovery tools for rapid entity and evidence extraction, relationship discovery, and semantic analysis. In multi-agency and multi-level security environments, where extremely large-scale datasets exist, the need for associating relevant data to certain relationships to a mission in hand or handling an unexpected event is ever more critical and challenging.

The associative model of data is an alternative data model for database system design. Other data models, such as the relational model and the object data model, are record-based, whereas in an associative database management system, data and metadata (data about data) are stored as items and links, which supports enhanced discovery by making connections not easily available from traditional database designs. Associative databases deliver enhanced discovery while minimizing CPU cycles, so performance on large datasets can be an important issue.

At this time there are no known open-source and nonproprietary associative database products available; expensive commercial hardware and software products are available for use but these are based on proprietary design and implementations. However, a number of fundamental enabling technologies contribute to the essential core of associative database design, including graph theory, resource description framework (RDF) triple storage and processing, attribute-value systems and entity-relationship modeling. Many of these supporting technologies have commercial big data applicability and cloud-computing open-source implementations are often available. It is envisioned that this effort will result in an associative database capability based upon both current state-of-the-art technologies and new methodologies and approaches that can be integrated in innovative ways to deliver enhanced capability across very large datasets in a multi-level secure environment.

This effort will produce an open-source and nonproprietary associative database prototype architecture that is scalable across multipetabyte datasets and incorporates security features applicable to deployment and suitable for a multi-agency, multi-level secure environment.

PHASE I: Investigate and evaluate the existing associative database published papers, techniques, and products. Identify the most promising approaches and perform tradeoff studies amongst those approaches for associative database concepts. Determine the technical feasibility for the creation of an open source and nonproprietary associative database architecture that would be scalable to petabyte stores, providing the required multi-level security access features, compatible with DoD Cloud data structures, and allowing for the plug and play of proprietary code. Through modeling and simulation determine the technical feasibility of the most promising candidate associative database design concept. Verify and validate the performance through implementation and demonstration of the basic proof-of-concept prototype that supports enhanced discovery on petabyte datasets suitable for multi-level security access features. Then document and report the results.

PHASE II: Based on the trade-off studies and analysis performed in Phase I amongst the competing concepts and approaches, determine the best possible approach to design and build a well-defined deliverable prototype system for the associative database and the ability to link to a cloud computing environment relevant to a mission of interest.

Implement approaches for exploring the potential automation of the extraction of relationship (triple store – entity – relationship - entity) data that could be facilitated through the use of language processing techniques on entity text strings identified through the use of associative processes. The final design must be robust to noisy data and scalable. The final prototype system also needs to be tested and performance demonstrated and validated at a government facility to show that the prototype scales to a multi-petabyte datasets across hundreds of nodes. The final report must include a detailed design of the system, technical documentation, and user manuals.

PHASE III: Incorporate the associative database technology in an existing or planned operational test environment at a designated Naval or a Joint Interagency Intelligence Analysis Center. Prepare plans to conduct numerous experiments with wide-ranging security and policy restricted requirements. Demonstrate the associative database system working in a cloud computing environment for a mission of interest and showing scalability to Petabyte stores. Demonstrate the automation of the extraction of relationship (triple store – entity – relationship - entity) data within a cloud enabled workflow using associative data processes. Collect performance data from field experiments to validate improved performance, scalability, and robustness of the system under extreme conditions that will include uncertain/incomplete intelligence data streaming in real-time. It is expected that the outcomes from each experiment will be measured and contrasted against the performance of the existing state-of-the-art database systems in place at the facilities. Prepare plans to garner acceptance and commitment from the multitudes of end-users, which requires tight coordination and collaboration with the analysts, operational communities, and subject matter experts. The associative database technology will pull-in sensitive data from various data sources, therefore the end-user commitment for implementation, training, and maintenance of the technology is required. Develop guidelines and documentation for transition including underlying applications, implementation procedures, and maintenance.

PRIVATE SECTOR COMMERCIAL POTENTIAL/DUAL-USE APPLICATIONS: This technology has broad applications for knowledge management and relationship extraction in both government and private sectors. In government it has numerous applications in military, intelligence communities, law-enforcement, homeland security, state and local governments to deal with asymmetric threats, deploy first responders, crisis management planning, and humanitarian aid response. The technology is equally compelling in commercial sector applications as it provides an environment to rapidly infer relationship and connect the right consumers to appropriate suppliers for wide-ranging services. In essence the associative database system enables rapid understanding of highly complex events and situations by "connecting the dots" in an environment that involves high data volume and quick response.

REFERENCES:

1. Williams, S., The Associative Model of Data, Lazysoft, 272p, www.sentences.com/docs/amd.pdf
2. Homan, Joseph V., Kovacs, Paul J., A Comparison of the Relational Database Model and the Associative Database Model, http://iacis.org/iis/2009/P2009_1301.pdf
3. The Associative Database in Support of Lean and Agile Database Design, <http://www.decisionsciences.org/Proceedings/DSI2008/docs/354-9596.pdf>

KEYWORDS: Associative Database, graph theory, resource description framework (RDF), triple stores, Petabyte, entity-relationship