

# Cogito



## Relationship Analytics™

Discover infinitely intricate patterns and distant relationships across your integrated data stores – in seconds

### In this issue...

- [Social Networks and Cogito Knowledge Center 2.3](#)
- [Finding a Needle in a Haystack](#)
- [A Better Defense](#)

### Quick Links

- [Our Web Site](#)
- [Our Products](#)
- [News and Events](#)
- [Contact Us](#)

### Join our mailing list!

## Rapid Report

*The latest news from Cogito*

October 2006

# Cogito

As relationship analytics and social networking continue to become more timely and topical Cogito continues to drive more functionality and capability in our products and services. In the October issue of our

news letter, Coleman shares with us a bit about version 2.3 of the Knowledge Center and Willy highlights the issues surrounding the intelligence and security arena and how critical Relationship Analytics are to finding a needle in a hay stack.

Please feel free to offer feedback, it is always good to hear from our customers and partners.

## Social Networks and Cogito Knowledge Center 2.3

*Coleman Barney, CEO*



Social networking and social-networking analytics seem to be the rage in the news and on the blogs. For Cogito, it's been invigorating to see the products we strive to create go from being sophisticated relationship analytics products to being "hip" and topical. It's like being a geeky scientist and all of a sudden lab coats and safety glasses are cool.

We've made significant progress with our Cogito Knowledge Center and invaluable help has come from social networking customers, which we appreciate. Next up by the end of this year will be v2.3 which will allow customers to better scale our solution for use online, managing hundreds of thousands of requests and calculations per day and customers will better be able to fuse data together and perform scheduled updates on their graphs.

Here's a summary of the feature set coming up in v2.3.

- **Modular, scalable server architecture:** The new graph engine monitor increases the number of concurrent Knowledge Center users and the practical scale of the graph store through advanced session control, load balancing capabilities, and the ability to distribute the Knowledge Center backend across multiple servers.
- **More powerful source data connections:** The Data Integration Broker provides live connections to data sources such as log files and databases which can now drive continuous, real-time updates to the graph store.
- **New Administration and Reporting tools:** The web- based Monitor Administrator Interface provides real time health and performance metrics

of mission critical server modules including the monitor and graph engine.

In addition to these and other enhancements, v2.3 will include a complete SDK for custom applications and solutions integration. Stay tuned for more launch details and timing to be available in October.

### **Finding a Needle in a Haystack**

*William Donahoo*

The world now operates differently; the Internet age is real, technology powers many facets of our lives and with it comes the ability to generate massive amounts of text, data and information.

The challenge (or is it opportunity?) is sifting through and leveraging this data and information to form actionable intelligence and glean the knowledge out of the massive amounts of information, or as we used to say “find the needle in the haystack.”

Actually, finding the needle in a haystack is an understatement. To give you an idea of how big this haystack is, two weeks of mobile phone call records alone can easily fill a database to four terabytes, RFID records are said to generate millions of terabytes of tracking information and the Internet, the fastest growing medium of all time, has been estimated to have more than half a billion terabytes of web information.

Current database and data mining capabilities are limited in their abilities to work with large amounts of data. While effective at housing and providing ready access to raw data, the ability to find links, identify patterns and highlight distant relationships are not possible – and to perform undirected discovery and hypothesis generation are only wishes.

We often read about the need to rethink how to deal with the data explosion that is happening around us and how RDBMS solutions are not the end-all-be-all we thought they once were. References to the relational schemas and how the schema diversity explosion is causing the pure relational model to collapse under its own weight are more frequent and clear in our understanding.

Driven initially by the demands of national intelligence and security organizations and quickly followed by other federal agencies, commercial businesses, Internet communities, online advertising and advanced market segmentation, it is clear that the need to mine data continues to accelerate. The demand for better ways to leverage information and improve the future by understanding the past are growing and new approaches and methods to data mining are required.

Fortunately, new research in graph-based math theory and graph-based data mining methodologies are providing the solutions required to find the needles in the haystacks.

Now analysts, market researchers, investigators and detectives have the ability to rapidly identify non-obvious relationships, hone in on replicated patterns and define how entities like people, events and actives are linked together. This new capability aids in defining behavior, showing roles and inferring possible future activities and actions.

By maintaining a natural index of all relevant entities in a graph database with the ability to relate, track, measure and visualize them over multiple degrees - gives data analysts the flexibility and speed needed to quickly narrow the scope of an investigation, find the non-obvious relationships and provide new and actionable intelligence to solve crimes and better enforce the law.

Crimes are often preceded by a number of suspicious or illegal activities—a stolen

vehicle or equipment, surveillance of a targeted area, change in daily routines, increased (often indirect) communication, new training sessions, etc. While any single one of these isolated activities is not alarming, put together they can identify a potential crime or terrorist activity that, with advance knowledge, can be interrupted and avoided.

An example close to home makes the point a bit more clear. Imagine the following activities:

- Sept. 1994 - A non-farmer buys one ton ammonium nitrate
- Oct. 1994 - The same non-farmer buys another ton of ammonium nitrate
- Oct. 1994 - The same month the same person purchases \$3,000 worth of nitromethane racing fuel
- Dec. 1994 - The Alfred P. Murrah Federal Building is cased
- Apr. 14, 1995 - Timothy McVeigh buys a used car and checks into a hotel in Kansas
- Apr. 16, 1995 - Timothy McVeigh has his friend, Terry Nichols, drive him back to Oklahoma
- Apr. 17, 1995 - Timothy McVeigh (using the alias Robert D. Kling) rents a 20-foot Ryder truck
- Apr. 19, 1995 9:02 a.m. - A bomb rips through the Murrah Federal Building, killing 168 people including 19 children and injuring hundreds of others.
- Apr. 19, 1995 10:30 a.m. - Timothy McVeigh is arrested 90 minutes later on firearms charges after a traffic stop near Billings, Okla.
- April 21, 1995 - Shortly before being released, McVeigh is recognized as a bombing suspect and charged in the bombing.

Had public safety investigators been able to track or trace the suspicious purchases of bomb-making materials and interest in the Federal Building they could have tagged McVeigh as a possible domestic terrorist, drilled down on the common relationships and charged him with "threatening to blow up the Alfred P. Murrah Federal Building" before any damage was done. While fortunately he was caught, the sad part was it was a coincidence and AFTER 168 innocent people were killed.

Another example is created in the graphic below. Here it shows how a relationship was identified between a U-Haul truck that was not returned and a large fertilizer purchase. Here it was found that the individual who rented the truck had been incarcerated with the individual who purchased the fertilizer. In looking at other connections, it was found that they had both been cited in separate instances at the Cooley Dam. Now authorities are able to investigate a potential threat to the dam involving these two individuals and a truck bomb.

As this example illustrates, it is not often that a conspirator carries out a criminal activity on their own. As such, it is not readily apparent with traditional data mining technology how isolated events may be connected to one another.

We now have more information than we have ever had before, but the vastness of that information has made it difficult to identify relationships and patterns that may have security implications. With graph-based relationship analytics data analysts and security directors can finally fully leverage their data. And while haystack may continue to grow, but the needle just became much easier to find.

**A Better Defense**

*Shoring Up Security Means More Than High Technology*

*Excerpt from article appears in the October 2006 issue of Scientific American*

One time-tested method for addressing the front end is surveillance and infiltration, as in the case of the foiled London plot (although some experts question the timing of the arrests, arguing that further observation might have revealed additional suspects or terrorist networks). Monitoring Web sites and data mining are also critical. Once intelligence agents have collected interesting data sets, "they need to be able to connect the dots," says William Donahoo, vice president at analytical software developer Cogito, which counts the National Security Agency among its licensees. Cogito software works by linking so-called nodes--people, places or things--via their various arcs, or relationships, and then generating hypotheses based on elements such as the number of relationships a given node has. Of course, the mining is only as good as the data. "It's not meant to automatically identify bad people and have them arrested," Donahoo adds. "We're trying to help analysts do their job better."

[Read full report](#)

