EMBRACING DATA ANALYTICS:

COMMON CHALLENGES & HOW TO OVERCOME THEM

USING ANALYTICS TO IMPACT THE MISSION VS. OBSERVE THE MISSION

An interview with Romi Doshi, Senior Solutions Manager at MarkLogic

The amount of data that agencies need to find and parse through at any given time can be daunting. In addition to analyzing mountains of information, analysts need to determine how the insights they glean can improve mission operations.

Before using data analytics to improve their missions, agencies should review their enterprise data management model. Does it support transitioning from simply observing the mission, by using analytics to determine what has happened, to actually impacting the mission by dynamically updating information and actions as events occur?

"Before agencies even get to analyze their environment, they are bombarded with data from many different sources," said Romi Doshi, Senior Solutions Manager at MarkLogic, an industry leader in enterprise database platforms. "It is overwhelming, and they face many challenges around data management and integration with legacy systems. Data today resides in so many different silos, so it's extremely difficult to get a 360-degree view of all information."

For example, let's say your agency is tasked with identifying and stopping illicit activities. Today, many analysts and decision-makers are pulling information from different systems. Reviewing this data is a costly manual effort because they have to go from one system to another to compile the information they need. But the work isn't done just yet because the bits and pieces of data you've gathered need to be aggregated and correlated in another system for further analysis. Many agencies use relational databases or data warehouses to integrate such silos. This is costly and greatly delays decision-making, Doshi explained.

Other agencies use complex ETL (Extract, Transform and Load) infrastructure to move data across multiple silos.

"The problems with ETL are its operational costs and lack of flexibility," Doshi said. According to one study, organizations spend about 80 percent of their time just wrangling data to perform ETL, which is a huge chunk of the enterprise data management budget.

What happens when new data sources are discovered and need to be added? Agencies would first have to understand the data and the metadata that resides in the new sources, prioritize which sources are the most important, redesign schemas, perform the ETL processes, rebuild the data warehouse, rinse and repeat.

And unfortunately, after all is said and done, you're still left with static data in your data warehouse or across silos. Think back to the earlier example of the organization combatting illicit activities. Information to stay on top of this task requires logins to multiple data sources or costly ETL to map to a data warehouse. By the time an analyst completes a report, the information is likely outdated.

A data warehouse is read-only, and it's meant to just be queried, Doshi noted. "At MarkLogic, we call that observing the mission. So you're observing what transactions have happened in the past and trying to gain intelligence from the data warehouse."

The contrast is actually observing and impacting the mission by using an operational and transactional database that allows you to feed new data sources and update information back into the system in — real time. That's where MarkLogic's Operational Data Hub solution has been an asset to agencies. It allows them to expand beyond a read-only data warehouse or point-to-point ETL-driven data integration.

An operational data hub built on the MarkLogic Enterprise NoSQL multi-model database platform can ingest information from the hundreds or thousands of data silos across departments, without requiring agencies to create a schema upfront. As agencies load new data into the system, analysts can not only immediately search across all the data, but also edit and update information as needed to make new connections between the data and write back to the database, while having traceability to the original sources.

Doshi equated the ease of MarkLogic's operational data hub to something we're all familiar with: Google's search engine. "If I create a new website, Google doesn't ask me to fill out my information in a specific form in order for it to be recognizable by their search engine," she said. "Google just scours the internet for new web pages and content, and it automatically shows up in search results. That's how MarkLogic works. You don't have to fit data into a specific mold like a schema, you just have to load data as is into MarkLogic, where it's indexed and searchable immediately. Over time you can enrich data to transform it into information and organizational knowledge."

For agencies, these capabilities mean less time spent prepping for analysis and more time spent extracting valuable insights.

govloop