HOW UNIFIED DATA ARCHITECTURE CAN REVOLUTIONIZE ANALYTICS IN GOVERNMENT

TERADATA

EXECUTIVE SUMMARY

Data is a highly sought-after commodity, but simply amassing information isn't enough. To be useful, the data must be actionable, and must be presented in a way that makes it meaningful. And it's that piece of the puzzle that's hanging up government data use.

Data – including big data, which is so large or so rapidly produced that traditional technology can't process it – is not new. Public and private entities have been generating it for decades. What is new, however, are the kinds of data emerging and the technology that makes data more readily functional. And as is the case with any new technology, there's a learning curve.

Understanding how to use data effectively is important, especially for government agencies. It can enable them to meet their mission objectives faster, better engage citizens, and cut costs. To learn how government agencies can go from reams of context-less numbers to usable statistics, we spoke with Alan Ford, Director of Pre-sales Consulting for Teradata Government Systems.

Ford said that the best way to make sense of data is not to look at it in silos, but rather to integrate it in a single unified data architecture (UDA). That way, agency officials can see how seemingly disparate datasets actually work together to produce a more complete picture.

In this industry perspective, we look at:

▷ How data is changing and the resultant challenges associated with making sense of new information.

▷ Case studies of how UDAs are helping the transportation industry.

▷ The benefits of having a UDA and how to overcome barriers to attaining valuable data analytics.

There's no doubt that data can be daunting. But using it effectively is easier than many government workers likely think. You just need the right tools to properly collect, process, and analyze the information. These tools will help you answer the question, "What would you do if you knew?"

NEW DATASETS, NEW ANALYTICS

Data is growing at unprecedented rates. Research firm IDC expects it to grow to 40 zettabytes (one zettabyte is equivalent to one billion terabytes) by 2020, a 50-fold growth from 2010. What's more, the type of data that's being generated is changing. In the past decade, organizations have started generating and encountering more semi-structured and unstructured data, which has no pre-defined data model or organization. This information is coming from sources such as e-mail and sensors. In fact, 90 percent of all data created in the next decade will be unstructured, market research firm IDC found.

"What you're going to see is a literal explosion of data — semi-structured or unstructured — that is going to be available and need a powerful platform on which to conduct analytics," Teradata's Alan Ford said. "As data volume grows exponentially over time, it's going to become more important to have ways to exploit it quickly."

In addition to the volume and velocity at which data is being produced, these varying formats add another layer of difficulty to making use of all the information. The government recognized that when Congress passed the Digital Accountabil-

ity and Transparency Act (DATA) of 2014, which Teradata heavily supported. The act requires all government agencies to produce data in a standard format so that financial, budget and spending data from myriad agencies can be consumed by anyone to determine how taxpayer money is being spent.

Putting everyone on equal footing will go a long way toward creating a data-driven government in which decisions are based on facts and continuously updated statistical analyses which, in turn, will effect more educated choices.

WHAT IS A UNIFIED DATA ARCHITECTURE?

With all this change, existing analytics no longer make the cut. This is where a UDA makes all the difference. But first, what is a UDA?

"Generically, a unified data architecture is any architecture which seeks to combine analytics capabilities across a plethora of data types," Ford said.

To create its UDA, Teradata added the Aster Discovery Platform and a data platform to work with the Teradata Integrated Data Warehouse. The Aster Discovery Platform uses a style of analytics known as MapReduce, which until now required a highly specific skill set to use, and built

it into a Structured Query Language (SQL) engine. (Teradata acquired Aster Data Systems and its SQL-MapReduce-based platform in 2011.)

"Now, instead of requiring an army of Ph.D. programmers to write MapReduce code, more than 100 popular MapReduce algorithms have been built into the Aster SQL engine. Folks who are Structured Query Language programmers, who are the majority of end analytics users, can access those algorithms fairly easily just by plugging in parameters to them rather than writing actual code," Ford said.

"The data platform can be either Hadoop or a Teradata Integrated Big Data Platform. The use of multiple platforms allows an organization to analyze its data regardless of type and location of the data in the UDA," Ford said.

"Enterprise data warehouse users, before they had an enterprise data warehouse, were used to analyzing information in a siloed fashion," he said. "For example, a bank might keep all of its customer data in one silo and its financial data in another silo. An integrated data warehouse environment seeks to load all of those data in a single platform for analysis across all of the different data subject areas."

But as we mentioned, less-structured or unstructured data are the beasts to tackle today, and traditional platforms won't be able to load or easily host semi-structured data such as JavaScript Object Notation (JSON), a data-interchange format based on the JavaScript scripting language. To that end, Teradata has built a JSON data type into its architecture, which means that data requires minimal processing up front to be loaded onto the database and be available for analysis.

The Aster Discovery Platform allows the hosting of these different types of data and reduces the overhead for processing data into load-ready form.

"This means that data will already come with the necessary information that allows it to be loaded just as it is received and still be analyzed by the various analytic algorithms that are built into the Discovery Platform," Ford said.

WHY YOU NEED A UNIFIED DATA ARCHITECTURE

▷ A unified data architecture provides government with the ability to ask questions of its data in order to advance future questions.

▷ A UDA allows organizations to capture, deploy, support, manage, and seamlessly access all its data.

▷ The biggest opportunities to government agencies of data analytics span the efficient service delivery to constituents, effective resource utilization, and cost containment.

▷ Agencies that don't have or don't plan to develop a data architecture risk future inefficiencies in the areas of cybersecurity, fraud management, tax management, DATA Act compliance, transportation analytics and much more.

THE UNIFIED DATA ARCHITECTURE IN ACTION

To see what a UDA can do to elevate business performance, look no further than the transportation industry. The airlines were one of the first to tackle the concept of growing data volumes. They were the first to demonstrate that with bigger data, solutions aren't found by just adding more data to existing business approaches, but by finding new ways to embed new data and analytics into most core processes, said Peeter Kivestu, Senior Industry Consultant Transportation at Teradata, in a recent GovLoop online training.

Facing competition from emerging low-cost airlines in the 1980s, established airlines began using booking data for more than just selling seats and checking in passengers. Even though keeping booking data history was considered unimaginably big data at the time, they created new analytics processes for managing pricing and seat availability, and they even invented new value propositions. By capturing and using booking data history, they created the modern-day, mileage-based frequent flyer programs. By simply taking "waste data," such as a completed booking, packaging it up for individual travelers and sending them monthly statements, they showed how close they were to accumulating enough miles for a free trip.

Applying that kind of thinking to the public sector is what Singapore Land Transportation Authority (LTA), which manages multiple modes of transportation, did with its substantial data collections. They recognized that while they used vast quantities of detail data every day to handle riders, its information technology systems were not designed for analytics, leaving officials with only three months of data online at a given time. That type of limited historical data set was not nearly enough for meaningful analyses, he said.

With that in mind, LTA built a customer-centric analytic solution by integrating anonymized data from disparate systems to get a customer's perspective on end-to-end travel times across all modes. LTA recognized there was a need to provide for high-volume data crunching under many conditions, ranging from statistical insights to regulatory reporting and predictive analytics, Kivestu said.

The benefit was a single source of people-centered insights that helped LTA focus on making public transport a choice mode.

For instance, there are lessons here for opportunities in public sector highway user experience management. Kivestu added, if you want to know why speed on highways is variable, you need data on volume and traffic incidents and weather to filter out the high-value opportunities and find the insightful information.

"It is through the study of detailed data that new insights emerge," Kivestu said. "Finding actionable opportunities depends on being able to connect the dots. Oftentimes, in

order to use one piece of data, you need other context data."

In Virginia, transportation officials have turned to variable-toll roads that adjust pricing based on statistics such as time of day (rush hour vs. the middle of the night, for example), volume and speed.

"It all goes back to theories of economics," Ford said. "If you're on a very busy road and the traffic flow has degraded and slowed, if there is an alternative that allows you to pay to get out of this situation, you're probably going to want to pay more for it when the roads are busy as opposed to when they are not busy."

It sounds simple, maybe even intuitive, but sophisticated analytics are necessary to determine such decisions as what the tolls should be, when they should be changed and what the maximum price is. As roadways get smarter and can determine the amount of traffic flow via sensors both in the roads and

individual vehicles, that information must be analyzed in real time to adjust rates accordingly, Ford said.

A good transportation data system needs three main data assets, Kivestu said. These are:

▷ Rich internal core data.

▷ External data that gives context to the internal information.

▷ User-generated data, such as that from social media and mobile apps.

Using separate systems to analyze each data set would do a disservice to transportation consumers. Without that contextual angle, the information might be interesting but perhaps not useful. And trying to piece them all together after the fact would mean long hours for public sector employees at the cost of taxpayer dollars.

"IT IS THROUGH THE STUDY OF DETAILED DATA THAT NEW INSIGHTS EMERGE. … FINDING ACTIONABLE OPPORTUNITIES DEPENDS ON BEING ABLE TO CONNECT THE DOTS."

KNOCKING DOWN BARRIERS

Looking at the transportation examples, we see how a UDA gives agencies insight into ways they can provide better customer service and maximize their agency dollars. But extrapolating further, we can envision use cases in cybersecurity, fraud management, legal compliance and more. The key differentiator of a UDA compared to other data analytics solutions is having all the data available in one place. The opportunities, like the questions that will emerge from the data, are endless.

Still, many government agencies continue to tread carefully when it comes to analytics. What's holding them back?

Ford highlighted two barriers: expertise and cost.

"The biggest barrier is probably the expertise to be able to exploit [data], to understand and to manipulate it," Ford said. However, the Discovery Platform's SQL interface has eliminated that obstacle.

A unified architecture also makes analytics less expensive. Previous efforts to analyze any amount of data required large amounts of storage, and storage — especially storage on technology that is structured such as a traditional database platform — is much more expensive than technology that can handle data in any format.

"We needed a platform that could host data in a relatively inexpensive way, and that platform is turning out to be the Hadoop platform," Ford said. "Now that Hadoop is widely available with both commercial and open source versions, it really just requires the ability to set up the file system and a set of relatively inexpensive disks to host it."

Additionally, agencies can use UDA platforms virtually if the data sizes are small enough, which also lowers costs. Agency IT managers who might be hesitant to purchase hardware nodes and disk arrays can start this way and move to a stand-alone system when data needs demand it.

CONCLUSION

The move toward an analytic culture full of data-driven enterprises means that government is starting to ask more questions of its data – and questions beget questions. In this environment, agencies need a discovery-oriented architecture that enables any query by any user at any time. That's what a UDA provides. Moreover, a UDA also removes barriers to data analytics and lowers data collection, storage and analytics costs while helping public-sector organizations better and more quickly meet constituents' needs.

ADDITIONAL RESOURCES:

▷ White Paper - Teradata Unified Data Architecture in Action

▷ White Paper - Teradata Unified Data Architecture: Integrated Data, Strategic Insight, Business Action

▷ Webinar: Driving new Insights Through Next-Gen Transportation Analytics

▷ Brochure – Keep Transportation Moving with Analytics

ABOUT TERADATA

Teradata helps government agencies get more value from data than any other company. Our big data analytic solutions, integrated marketing applications, and team of experts can help your organization achieve actionable insights with data. Teradata helps organizations leverage all of their data so they can know more about their citizens and agency outcomes and do more of what's really important. With more than 10,000 professionals in 43 countries, Teradata serves top companies across government, consumer goods, financial services, healthcare, automotive, communications, travel, hospitality, and more. A future-focused company, Teradata is recognized by media and industry analysts for technological excellence, sustainability, ethics, and business value.

VISIT TERADATA.COM/GOVERNMENT.

ABOUT GOVLOOP

GovLoop's mission is to "connect government to improve government." We aim to inspire public sector professionals by serving as the knowledge network for government. GovLoop connects more than 150,000 members, fostering cross-government collaboration, solving common problems and advancing government careers. GovLoop is headquartered in Washington, D.C. with a team of dedicated professionals who share a commitment to connect and improve government.

For more information about this report, please reach out to info@govloop.com.

1101 15th St NW, Suite 900
Washington, DC 20005

Phone: (202) 407-7421 | Fax: (202) 407-7501

www.govloop.com
@GovLoop

govloop