book

NOVEMBER 2016

BIG DATA, ANALYTICS, AND THE CLOUD:

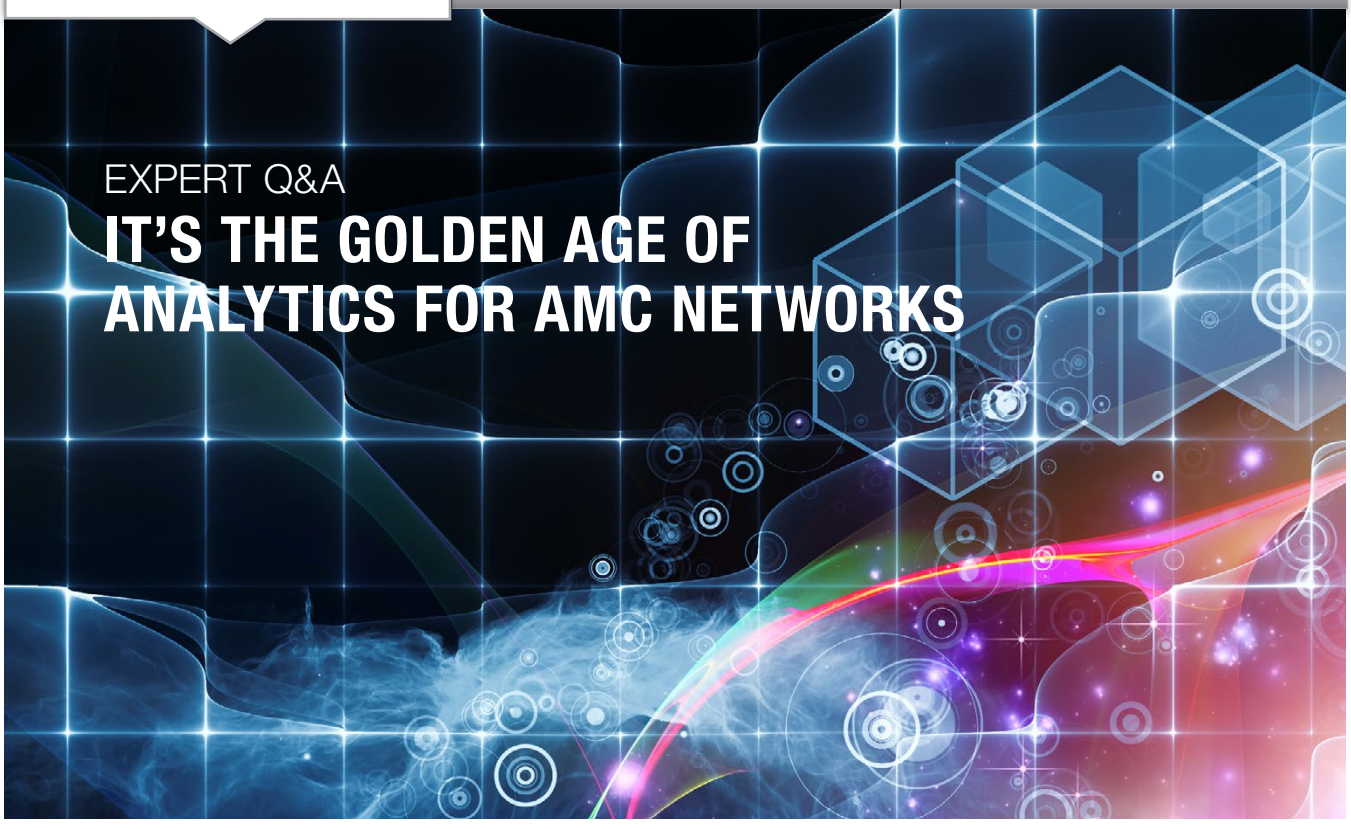
STRATEGIES FOR SUCCESS

Sponsored by:

IBM

tdwi

Transforming Data With Intelligence™



EXPERT Q&A

IT'S THE GOLDEN AGE OF ANALYTICS FOR AMC NETWORKS

French scholar Alex Picard has argued that, since the dawn of the 21st century, we have entered a "Golden Age of Television." He bases this on three primary elements—first, an improvement in both visual aesthetics and storytelling; second, an overall homogeneity between cable and network series; and third, tremendous popular success.

However, he is surely overlooking a fourth key element—big data analytics, the cloud, and the role they play in matching viewers with the content they find most appealing.

TDWI spoke with Vitaly Tsivin, senior VP of business intelligence for AMC Networks, to understand more about the kinds of analytics AMC performs, and why they made the move to the hybrid cloud.

TDWI: What kind of data do you collect and how do you use it?

Vitaly Tsivin: We utilize a lot of data. For instance, we work with census data. There are 250 million [adults] in the United States with, I believe, 3,000+ [census] characteristics each. We utilize panel data of 100,000 households and what they watch, minute by minute, for the past six years. We also collect social media data and other digital data, as well as survey data. That allows us to work with hundreds of trillions of rows of data.

It is about 250 terabytes of data.

We use our first-party data to identify seeds. For example, when people sign up for newsletters for a given show, we then identify those people in the census world and use the data sets as a seed to build statistical lookalike models. We then rate everyone in the U.S. as to how closely they resemble the group of people who show extreme affinity for our product and we provide direct-to-consumer marketing to those people.

How do you manage and analyze all of that data?

For longest time we were satisfied with our Netezza appliance from IBM, along with Cognos and SPSS and other tools such as MDM—all driven by the appliance box. The drawback of the appliance, however, is that it is not scalable. You really can't even say it's a disadvantage because that's the reason it's called an "appliance". If you run out of space, memory, and so on, you can't farm it out. Eventually we hit the point that between the algorithms we were running on the appliance and more users wanting to use it, we hit a congestion situation.

So what did you do?

We started looking for a scalable solution. We asked, "Does it make sense to put some of this in the cloud?" We analyzed the market and decided to stay with IBM and use the BlueMix cloud in a hybrid fashion. Because of the nature of what we do, the "mother ship" is the appliance. One of the reasons we chose IBM was that they provide proprietary, extremely fast connectors to their appliances. Additionally, IBM has managed to combine the depth of a mature database—that's a database that provides complex support for complex procedures—with in-memory processing. It allows for flexibility in a very rich environment that is also very fast. Add scalability to that and suddenly you see an almost perfect product.

How did this impact the way you do your analytics?

In the world of big data, you're always in a position to aggregate, to create various snapshots and dumps of the data so that the user has the ability to get the data instantaneously. Rather than do that within the appliance itself, we are using this very fast connectivity to spin up data and analyze it.

Now, it would be an undertaking to run the lookalike algorithm I mentioned in the cloud. The best place to run that algorithm, which takes 10 to 30 minutes of real machine time, is on the appliance. What happens then is that when someone executes this algorithm or calls for an execution, the call is being made from the cloud and into the appliance. It's an asynchronous call. You can continue working on the system you are on, doing other things, and meanwhile you'll get an update of all the scheduled things that you've asked the system to do for you. The signal has gone to the appliance that is sitting inside the firewall and it uses the appliance's resources to do the calculations. Once the algorithm has finished the calculations and, say, the promotion plan is ready, it will be pushed back to the user via the VPN with a green flag, meaning the process is complete. When it comes back, it actually brings in whatever data it needs with it, so the heavy volume—the heavy processing—takes place in the appliance and only the result is being sent back to the cloud.

What has been the result of the hybrid cloud deployment?

The architecture allows us to onboard an almost unlimited number of users, which has allowed us to actually provide services to advertisers and partners outside of the company. That's been a big step for us.

ADOPTION PATTERNS FOR BIG DATA ANALYTICS IN THE CLOUD

By Fern Halper and David Stodder

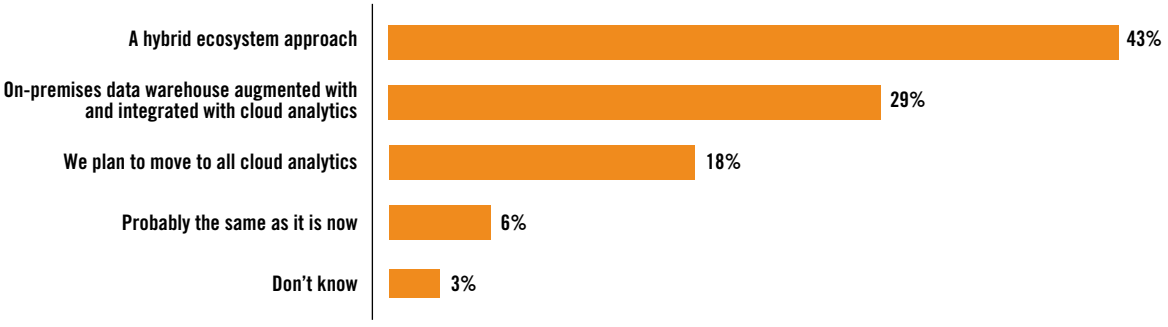
Today, many organizations are collecting more data than they can manage or analyze, even as they realize that big data and analytics can provide important strategic and competitive advantages. As part of this trend, the cloud is becoming an increasingly popular platform for BI and analytics.

As recently as three years ago, when TDWI asked respondents whether they were using or planning to use any kind of cloud for analytics, about 25 percent said that they would never use the cloud. Resistance is definitely diminishing, though, as organizations start to understand the benefits of the cloud for a range of business applications, including BI and analytics. For example, in the survey for our most recent Best Practices Report, only about 10 percent of respondents said they would never use the cloud for analytics.

Organizations like the scalable and flexible nature of the cloud. In fact, in our research, the top drivers for cloud analytics are "scalability" and "flexibility." Organizations like the ability of the cloud to scale on demand and the flexibility and simplicity of the cloud. Companies can deploy servers as needed and access analytics software quickly—which is especially appealing to those accustomed to long waits when ordering hardware and software. It is also appealing to those dealing with ever-increasing volumes of new data types. TDWI research shows that the cloud provides measurable value for organizations dealing with big data because of its elasticity and scalability.

In part due to these features, some organizations are using the cloud in their data warehouse modernization efforts. Data

Where do you see your analytics architecture in the next three to five years?



warehouse modernization is often referred to as *augmentation*, *automation*, *optimization*, or *modernization*, and often includes tools and platforms built to accommodate big data from new sources.[1] Often, organizations augment—but do not replace—their existing data warehouse when they add additional platforms and tools such as Apache Hadoop (with or without integrated MPP databases), appliances, newer kinds of cloud data warehouses, or other kinds of data management platforms. Some of these additions may be cloud-based so that the cloud becomes part of the modern data and analytics ecosystem—which creates what might be called a hybrid cloud implementation. This, then, also includes the integration tools and platforms necessary to integrate various cloud and on-premises environments.

Alternately, organizations might be using a big data platform—such as an appliance—but want to better manage their workloads. They may leave the appliance to do more data intensive analysis and use the cloud for other analytics activities.

Organizations are also beginning to collect data where it is created and analyze data where it is collected. A considerable amount of data is now being generated in the cloud—social media data, IoT data, and data from other cloud applications, to name just a few. Because it often makes sense to analyze data where it lives—what some in the industry refer to as "data gravity"—this may explain increasing acceptance of cloud BI and analytics for big data. We are seeing organizations doing the following:

- **Big data experimentation**. Cloud-based implementations also provide organizations with ways to explore new methods for managing and analyzing the huge volumes of data they are now generating. For instance, an organization may want to try Hadoop or Apache Spark but because it has not used these technologies before, it does not want to invest in the necessary infrastructure just to find out if they might be valuable. Hadoop-in-the-cloud is a good way to understand whether Hadoop is a good fit in general for the organization before investing heavily. The cloud also enables experimentation with different kinds of data management approaches, including data warehouses, data lakes, and open source technologies.
- **Analytics data reduction.** Organizations that are dealing with thousands of attributes across millions of rows of observations might use the cloud to initially determine what attributes are important for analysis. Those attributes that are viewed as important would then be sent to an on-premises platform for analysis.
- **Analytics sandboxes.** Other organizations use a public or private cloud as a sandbox to experiment with different algorithms or to build proofs of concept (POC). These

[1]For more information on data warehouse modernization, see the 2016 *TDWI Best Practices Report: Data Warehouse Modernization in the Age of Big Data Analytics*, online at tdwi.org/bpreports.

user-controlled spaces can use any kind of data and are often separated from production environments for power users or business analysts. For example, a company might use an analytics sandbox to experiment with predictive analytics to understand customer behavior and predict certain kinds of actions.

- **IoT analytics.** The Internet of Things (IoT) generates a lot of data, some of which is streaming. The cloud (public or private) is often a part of an IoT implementation. For instance, an insurance company might collect telematics data from vehicles that it then uploads to the cloud for driver risk analysis.

Ultimately, many organizations will make use of hybrid cloud environments for big data analytics. A hybrid cloud is one that includes the both public and private clouds, often with one or more touchpoints between them. A hybrid cloud may also include some sort of integration process to enable the cloud and on-premises environments to work together. As stated above, this may be part of a data warehouse modernization effort, or it may grow organically as an organization begins to leverage the cloud more frequently.

In a recent TDWI study, when asked about where they see their analytics architecture in the next three-to-five years, 28 percent said they would be using cloud analytics together with their on-premises data warehouse. An additional 43 percent believe that their analytics will utilize a hybrid approach, where the cloud forms one component of a bigger architectural strategy. These modern, hybrid data environments incorporate big data platforms, analytics, and Hadoop with traditional enterprise data, the BI/DW technology stack, and numerous IT data platforms—both old and new—to support analytics. This is where big data analytics is heading.

www.ibm.com

IBM Analytics offers a complete portfolio of big data and analytics solutions to fuel your cognitive business. Our solutions enable your organization to engage with data to answer the toughest business questions, uncover patterns, and pursue breakthrough ideas.

IBM Data Warehousing is a key set of offerings in this portfolio. These offerings enable a hybrid data architecture, with solutions to address a wide range of data types and workload deployment needs. IBM Analytics also offers solutions in data management, cloud data services, Watson Analytics, weather data, and industry solutions.

Transforming Data With Intelligence™

tdwi.org

TDWI is your source for in-depth education and research on all things data. For 20 years, TDWI has been helping data professionals get smarter so the companies they work for can innovate and grow faster. TDWI provides individuals and teams with comprehensive business and technical education and research that allow them to acquire the knowledge and skills they need, when and where they need them.

TDWI advances the art and science of realizing business value from data by providing an objective forum where industry experts, solution providers, and practitioners can explore and enhance data competencies, practices, and technologies.

TDWI offers six major conferences, topical seminars, onsite education, a worldwide membership program, business intelligence certification, live webinars, resource-filled publications, industry news, an in-depth research program, and a comprehensive website at tdwi.org.