

FEATURE

6 data analytics trends that will dominate 2018

As businesses transform into data-driven enterprises, data technologies and strategies need to start delivering value. Here are four data analytics trends to watch in the months ahead.

By Thor Olavsrud

Senior Writer, CIO

MAR 15, 2018 1:06 PM PT

Together with social, mobile and cloud, analytics and associated data technologies have emerged as core business disruptors in the digital age. As companies began the shift from being data-generating to data-powered organizations in 2017, data and analytics became the center of gravity for many enterprises. In 2018, these technologies need to start delivering value. Here are the approaches, roles and concerns that will drive data analytics strategies in the year ahead.

ADVERTISING



Data lakes will need to demonstrate business value or die

Data has been accumulating in the enterprise at a torrid pace for years. The internet of things (IoT) will only accelerate the creation of data as data sources move from web to mobile to machines.

"This has created a dire need to scale out data pipelines in a cost-effective way," says Guy Churchward, CEO of real-time streaming data platform provider DataTorrent.

[Learn the secrets of highly successful data analytics teams. | Beware the 12 myths of data analytics and the sure-fire ways organizations fail at data analytics. | Get the latest on data analytics by signing up for CIO newsletters.]

For many enterprises, buoyed by technologies like Apache Hadoop, the answer was to create data lakes — enterprise-wide data management platforms for storing all of an organization's data in native formats. Data lakes promised to break down information silos by providing a single data repository the entire organization could use for everything from business analytics to data mining. Raw and ungoverned, data lakes have been pitched as a big data catch-all and cure-all.

But while data lakes have proven successful for storing massive quantities of data, gaining actionable insights from that data has proven difficult.

"The data lake served companies fantastically well through the data 'at rest' and 'batch' era," Churchward says. "Back in 2015, it started to become clear this architecture was getting overused, but it's now become the Achilles heel for *real* real-time data analytics. Parking data first, then analyzing it immediately puts companies at a massive disadvantage. When it comes to gaining insights and taking actions as fast as compute can allow, companies relying on stale event data create a total eclipse on visibility, actions, and any possible immediate remediation. This is one area where 'good enough' will prove strategically fatal."

Monte Zweben, CEO of Splice Machine, agrees.

"The Hadoop era of disillusionment hits full stride, with many companies drowning in their data lakes, unable to get a ROI because of the complexity of duct-taping Hadoop-based compute engines," Zweben predicts for 2018.



SponsoredPost Sponsored by City of Hollywood
Grow Your Business in Florida's Hollywood

To survive 2018, data lakes will have to start proving their business value, says Ken Hoang, vice president of strategy and alliances at data catalog specialist Alation.

"The new dumping ground of data — data lakes — has gone through experimental deployments over the last few years, and will start to be shut down unless they prove that they can deliver value," Hoang says. "The hallmark for a successful data lake will be having an enterprise catalog that brings information discovery, AI, and information stewarding together to deliver new insights to the business."

However, Hoang doesn't believe all is lost for data lakes. He predicts data lakes and other large data hubs can find a new lease on life with what he calls "super hubs" that can deliver "context-as-a-service" via machine learning.

"Deployments of large data hubs over the last 25 years (e.g., data warehouses, master data management, data lakes, Salesforce and ERP) resulted in more data silos that are not easily understood, related, or shared," Hoang says. "A hub of hubs will bring the ability to relate assets across these hubs, enabling context-as-a-service. This, in turn, will drive more relevant and powerful predictive insights to enable faster and better operational business results."

SponsoredPost Sponsored by Chase
5 Reasons You Wouldn't Survive Without Your Phone





Ted Dunning, chief application architect for MapR, predicts a similar shift: With big data systems becoming a center of gravity in terms of storage, access and operations, businesses will look to build a global data fabric that will give comprehensive access to data from many sources and to computation for truly multi-tenant systems.

"We will see more and more businesses treat computation in terms of data flows rather than data that is just processed and landed in a database," Dunning says. "These data flows capture key business events and mirror business structure. A unified data fabric will be the foundation for building these large-scale flow-based systems."

Langley Eide, chief strategy officer of self-service data analytics specialist Alteryx, says IT won't be alone on the hook when it comes to making data lakes deliver value: Line-of-business (LOB) analysts and chief digital officers (CDOs) will also have to take responsibility in 2018.

"Most analysts have not taken advantage of the vast amount of unstructured resources like clickstream data, IoT data, log data, etc., that have flooded their data lakes — largely because it's difficult to do so," Eide says. "But truthfully, analysts aren't doing their job if they leave this data untouched. It's widely understood that many data lakes are underperforming assets — people don't know what's in there, how to access it, or how to create insights from the data. This reality will change in 2018, as more CDOs and enterprises want better ROI for their data lakes."

Eide predicts that 2018 will see analysts replacing "brute force" tools like Excel and SQL with more programmatic techniques and technologies, like data cataloging, to discover and get more value out of the data.

The CDO will come of age

As part of this new push to get better insights from data, Eide also predicts the CDO role will come into its own in 2018.

"Data is essentially the new oil, and the CDO is beginning to be recognized as the linchpin for tackling one of the most important problems in enterprises today: driving value from data," Eide says. "Often with a budget of less than \$10 million, one of the biggest challenges and opportunities for CDOs is making the much-touted self-service opportunity a reality by bringing corporate data assets closer to line-of-business users. In 2018, the CDOs that work to strike a balance between a centralized function and capabilities embedded in LOB will ultimately land the larger budgets."

Eide believes CDOs that enable resources, skills, and functionality to shift rapidly between centers of excellence and LOB will find the most success. For this, Eide says, agile platforms and methodologies are key.

Rise of the data curator?

Tomer Shiran, CEO and co-founder of analytics startup Dremio, a driving force behind the open source Apache Arrow project, predicts that enterprises will see the need for a new role: the data curator.

The data curator, Shiran says, sits between data consumers (analysts and data scientists who use tools like Tableau and Python to answer important questions with data) and data engineers (the people who move and transform data between systems using scripting languages, Spark, Hive, and MapReduce). To be successful, data curators must understand the meaning of the data as well as the technologies that are applied to the data.

"The data curator is responsible for understanding the types of analysis that need to be performed by different groups across the organization, what datasets are well suited for this work, and the steps involved in taking the data from its raw state to the shape and form needed for the job a data consumer will perform," Shiran says. "The data curator uses systems such as self-service data platforms to accelerate the end-to-end process of providing data consumers access to essential datasets without making endless copies of data."

Data governance strategies will be key themes for all C-level executives

The European Union's General Data Protection Regulation (GDPR) is set to go into effect on May 25, 2018, and it looms like a specter over the analytics field, though not all enterprises are prepared.

The GDPR will apply directly in all EU member states, and it radically changes how companies must seek consent to collect and process the data of EU citizens, explain lawyers from Morrison & Foerster's Global Privacy + Data Security Group: Miriam Wugmeister, Global Privacy co-chair; Lokke Moerel, European Privacy Expert; and John Carlin, Global Risk and Crisis Management chair (and former Assistant Attorney General for the U.S. Department of Justice's National Security Division).

"Companies that rely on consent for all their processing operations will no longer be able to do so, and will need other legal bases (i.e., contractual necessity and legitimate interest)," they explain. "Companies will need to implement a whole new ecosystem for notice and consents."

Even though GDPR fines are potentially massive — the administrative fines can be up to 20 million Euros or 4 percent of annual *global* turnover, whichever is highest — many enterprises, particularly in the U.S., are not prepared.

"When the Y2K boom came around, everyone was preparing for odds that they may or may not face," says Scott Gnau, CTO of Hortonworks. "Today, it seems that barely anyone is properly preparing for the GDPR being enforced in May 2018. Why not? We're currently in a phase where every organization is not only trying to deal for 'what's next,' but they're struggling to maintain and deal with issues that need solving *now*. Many organizations are likely relying on chief security officers to define the rules, systems, parameters, etc., to help their global system integrators figure out the best course of action. That is not a realistic expectation to put on one individual's role."

To enforce GDPR properly requires the C-suite be informed, prepared, and communicative with all facets of their organization, Gnau says. Organizations will need a better handle on the overall governance of their data assets. But large breaches, like the Equifax breach that came to light in 2017, means they will struggle to balance providing self-service access to data for employees while protecting that same data from prospective threats.

As a result, Gnau predicts data governance will be a focus point for all organizations in 2018.

"A key goal should be developing a system that balances democratization of data, access, self-service analytics, and regulation," Gnaou says. "The way we architect data safely going forward will have an impact on everyone — customers in the U.S. and overseas, the media, your partners, and more."

Zachary Bosin, director of solution marketing for multi-cloud data management specialist Veritas Technologies, predicts a U.S. company will be one of the first to be fined under the GDPR.

"Despite the impending deadline, only 31 percent of companies surveyed by Veritas worldwide believe they are GDPR-compliant," Bosin says. "Penalties for non-compliance are steep, and this regulation will impact every and any company that deals with EU citizens."

The proliferation of metadata management continues

It's not just the GDPR, of course. The data deluge keeps growing, and governments around the world are imposing new regulations as a result. Within organizations, teams have much greater access to data than ever before. This all adds up to increased importance of data governance, along with data quality, data integration and metadata management.

"Metadata management and ensuring data privacy for regulations such as GDPR joins earlier trends like AI and IoT, but the unexpected trend of 2018 will be the convergence of data management technologies," says Emily Washington, senior vice president of product management at data and analytics software provider Infogix. "Businesses are increasingly evaluating ways to streamline their overall technology stack if they want to successfully leverage big data and analytics to create a better customer experience, achieve business objectives, gain a competitive advantage, and, ultimately, become market leaders."

Extracting meaningful insights and increasing operational efficacy will require flexible, integrated tools that allow users to quickly ingest, prepare, analyze and govern data, Williams says. Metadata management, in particular, will be essential to supporting data governance, regulatory compliance, and data management demands in enterprise data environments.

Predictive analytics helps improve data quality

As data projects move into production, data quality is increasingly a concern. This is especially true as IoT opens the floodgates further. Infogix says 2018 will see organizations turning to machine learning algorithms to enhance data quality anomaly detection. By using historical patterns to predict future data quality outcomes, businesses can dynamically detect anomalies that might otherwise have gone unnoticed or might only have been found much later through manual intervention.

"As more data is generated through technologies like IoT, it becomes increasingly difficult to manage and leverage," Washington says. "Integrated self-service tools deliver an all-inclusive view of a business's data landscape to draw meaningful, timely conclusions. Full transparency into a business's data assets will be crucial for successful analytics initiatives, addressing data governance and privacy needs, monetizing data assets, and more as we move into 2018."

More on data analytics:

- 12 myths of data analytics debunked
- The secrets of highly successful data analytics teams

- 12 data science mistakes to avoid
- 10 hot data analytics trends — and 5 going cold
- 7 sure-fire ways to fail at data analytics
- 6 data analytics success stories: An inside look
- Demystifying data analytics: How to create business value with data
- Big data and data analytics certifications that will pay off

Next read this:

- *7 habits of highly effective digital leaders*
- *20 ways to kill your IT career (without knowing it)*
- *12 signs your strategic partnership has gone wrong*
- *6 most-dreaded IT projects*
- *8 secrets of effective change leaders*
- *8 early warning signs of IT disaster*
- *7 habits of highly effective digital transformations*
- *CIO playbook: 10 tips for leading IT in the digital era*
- *15 ways to advance your IT career*
- *7 steps to becoming a thought leader*

Thor Olavsrud covers IT security, big data, open source technology, Microsoft tools and servers for CIO.com.

Follow      

💡 **SUBSCRIBE! Get the best of CIO delivered to your email inbox.**