Syncsort's Fourth Annual Big Data Survey Uncovers Five Key Data Lake Trends for 2018

Reveals Enterprise Data Priorities, Use Cases and Opportunities for the Coming Year

Pearl River, NY – December 18, 2017 - Syncsort, a global leader in Big Iron to Big Data solutions, today announced the results from its fourth annual Big Data survey. Results revealed the top use cases and challenges faced by organizations as they progress their modern data architecture and data lake initiatives, and see significant benefits from Hadoop/Spark, with nearly 60% citing both increased productivity across the organization and improved efficiency to reduce costs as their biggest gains.

Compared with last year's survey, the most dramatic increase in reported benefits was higher revenue and accelerated growth, which 55% named as a benefit this year compared with 37% last year. And though organizations are using Big Data insights in more sophisticated ways to improve revenue and customer service, they continue to face some of the same challenges reported in past years, including keeping up with rapidly changing technologies and tools.

The survey also found that Hadoop and Spark, which had high interest but low adoption at the time Syncsort launched the survey in 2014, are now in production or test at 70% of responding organizations. Specifically, this year, more than 40% of respondents say they are in production with either Hadoop or Spark, and 30% say they are engaged in a proof of concept or pilot program.

Based on the survey results, the five key trends in Data Lake initiatives that organizations need to monitor in 2018 include:

1. The composition of the data lake shifts. Traditional sources remain most popular for filling the data lake. Relational Database Management Systems (RDBMS) were chosen as the top source at 69%, up from 61% last year, surpassing enterprise data warehouses (EDW) at 63%. But newer sources grew, with NoSQL databases identified by 46% of respondents compared to 35% last year. Cloud repositories are also gaining strength as a data source (accounting for 40%) as more organizations leverage the cloud as a deployment platform.

2. Legacy platforms continue to play a significant role. Data from legacy platforms (such as the mainframe and IBM i) also make major contributions to the data lake: over 97% of respondents with mainframe believe it's valuable to access and integrate mainframe data into the data lake for real-time analytics – a 27% increase over last year. Over 90% of those with IBM i believe it's valuable to access and integrate that data into the data lake, not surprising as leaving behind decades of valuable data stored on these systems would seriously hamper their companies' analytics.

3. Data quality and regulatory compliance challenges are top of mind. While the skills shortage had been ranked the top challenge for three consecutive years, this year it fell to number two, replaced by mounting concerns over improving the quality of data in the data lake. Indeed, 40% said data quality was a significant struggle for their organization, likely a result of expanded use of data lakes driving an emphasis on improving data quality. But the survey also showed not everyone is making the connection between data quality and ROI. 60% of Financial Services and Insurance professionals said ensuring data quality is a top priority, compared with just 40% of respondents from other industries.

Nearly 40% of respondents also cited meeting regulatory compliance mandates as a significant challenge – one that will only increase as GDPR and other compliance mandates become a pressing reality for executives. Therefore, the scope of data governance is projected to expand as organizations will place a higher priority on putting processes in place that allow them to understand what their data is and where it has been.

4. Data lakes will be kept fresher to support data use. This year, 71% of respondents named ETL as their most compelling data lake use case – a notable increase from last year, when it came in third at 53%. The second and third-highest use case involved analytics, with advanced/predictive analytics at 64% and real-time analytics at 60%. All of these use cases require fresh, up-to-date data; however, more than 75% of respondents report having difficulty keeping their data lake in sync with changing data sources, particularly when the source is hard to access, like mainframe.

5. Organizations will continue to invest in Big Data. Around 90% of organizations found that leveraging Hadoop and Spark and shifting away from legacy systems proved valuable not only in driving data insights, but also in saving money (21% reported freeing up mainframe resources reduced costs). As more companies discover the cost savings resulting from optimizing traditional platforms, more money will in turn be funneled into funding more Big Data projects.

"We are seeing increased adoption of data lake initiatives where organizations are very focused on governance of the data in the data lakes, increasing benefits through advanced analytics and machine learning and deployment of hybrid environments including cloud," said Tendü Yoğurtçu, CTO, Syncsort. "But those benefits can only be unlocked if organizations have access to enterprise data, can create trusted data sets and establish effective data governance practices. This propels them to a place where they can not only adapt to digital disruption, but take advantage of it so their businesses thrive."

[Click here to download the full report.](#)

Methodology: Syncsort polled nearly 200 respondents including data architects, IT managers, developers, business intelligence/data analysts and data scientists at organizations involved with interested in Hadoop and Spark. Participants represent a range of industries including financial services/insurance, healthcare, government, telecommunications, retail and more.

About Syncsort
Syncsort is a trusted enterprise software provider and the global leader in Big Iron to Big Data solutions. More than 6,000 organizations, including 84 of the Fortune 100, use the company's products to solve their most complex data management challenges, on premise and in the cloud. Syncsort helps customers optimize traditional data systems and deliver mission-critical data from these systems to next-generation analytic environments. Its Big Iron to Big Data portfolio now features the #1 high availability product for IBM i Power Systems, powerful cross-platform capacity management, best-in-class mainframe app and machine data access & integration, and market-leading data quality capabilities. Rediscover Syncsort a **[syncsort.com.](#)**

Media Contacts:

Stephanie Kacoyanis
SHIFT Communications
Tel: 617-779-1826
syncsort@shiftcomm.com

Michael Kornspan
Syncsort Incorporated
Senior Director, Corporate Communications
Tel: 201-930-8216
mkornspan@syncsort.com