EXECUTIVE OVERVIEW

PREVIEW

2017 Trends in Data Platforms and Analytics

OCT 2016

Matt Aslett, Research Director

Krishna Roy, Senior Analyst, Data Platforms and Analytics

Jason Stamper, Analyst, Data Management and Analytics

James Curtis, Senior Analyst, Data Platforms & Analytics

The data platforms and analytics sector has changed considerably in recent years, starting at the bottom up with the emergence of new data platforms. As those continue to emerge, we are witnessing greater impact at the data management and analytics layers as enterprises evolve their strategies to take greater advantage of the increased data processing and analytics capabilities available to them.

ABOUT 451 RESEARCH

451 Research is a preeminent information technology research and advisory company. With a core focus on technology innovation and market disruption, we provide essential insight for leaders of the digital economy. More than 100 analysts and consultants deliver that insight via syndicated research, advisory services and live events to more than 1,000 client organizations in North America, Europe and around the world. Founded in 2000 and headquartered in New York, 451 Research is a division of The 451 Group.

NEW YORK
20 West 37th Street
3rd Floor
New York, NY 10018
P 212-505-3030
F 212-505-2630

SAN FRANCISCO
140 Geary Street
9th Floor
San Francisco, CA 94108
P 415-989-1555
F 415-989-1558

LONDON
37-41 Gower Street
London, UK WC1E 6HH
P +44 (0)20 7299 7765
F +44 (0)20 7299 7799

BOSTON
75-101 Federal Street
5th Floor
Boston, MA 02110
P 617-261-0699
F 617-261-0688

ABOUT THE AUTHOR

MATT ASLETT

RESEARCH DIRECTOR

Matt Aslett is a Research Director for the Data Platforms and Analytics Channel at 451 Research. Matt has overall responsibility for the data platforms and analytics research coverage, which includes operational and analytic databases, Hadoop, grid/cache, stream processing, search-based data platforms, data integration, data quality, data management, analytics, machine learning and advanced analytics. Matt's own primary area of focus includes data management, reporting and analytics, and exploring how the various data platforms and analytics technology sectors are converging in the form of next-generation data platforms.

Key Findings

Machine learning and deep learning will play a bigger role in taking predictive analytics into the mainstream in 2017 and beyond.

Spark will overtake MapReduce as the predominant compute technology for analytic-based Hadoop use cases.

Self-service data preparation and analytics products are evolving to meet the needs of data stewards and IT professionals to manage data governance and deliver managed self-service data services.

Corporate performance management (CPM) cloud services will likely enter the mainstream in 2017, moving beyond adoption by small firms and enterprise departments.

Data grid providers will continue to expand their functionality and morph into stand-alone in-memory data fabric providers, driven by growing customer demand for in-memory data processing.

Executive Summary

INTRODUCTION

Much of the attention in the data platforms and analytics sector in recent years focused on the potential impact of emerging data platforms, and with good reason. The emergence of Hadoop and NoSQL databases has enabled enterprises to take advantage of the new economics of data to store and process data – typically structured and semi-structured – that was previously too expensive and/or complex to store in conventional analytic databases.

To truly generate value from the new economics of data, enterprises need to not only store that data at a cheaper rate than ever before, but also process and analyze it at greater volumes and greater frequency than was previously affordable to uncover new business insight.

451 Research has long described adoption of Hadoop and big data platforms as a process that begins with low-cost storage of previously ignored data, continues with the aggregation and integration of large data sets for exploration and experimentation, and results in new approaches to analytics that deliver new business opportunities.

A prime example of this last step is machine learning, which was already a major focus of innovation in 2016, and looks set to reach new heights in 2017. Machine learning is not new – after all, it was first defined in 1959 – but it has risen in prominence in recent years thanks to increasing compute power and data volumes, combined with enterprises' desire to move beyond historical reporting and analysis to embrace predictive, descriptive and even prescriptive analytics.

To put it another way, having already asked what happened and why it happened, enterprises are asking what will happen, what is happening now and how they can influence what happens next.

Descriptive and prescriptive analytics are evolving disciplines that will likely get further mention in reports such as this in the years to come, but for now we expect machine learning and its role in democratizing predictive analytics to be a major focus in 2017. Once again, predictive analytics is nothing new. What we anticipate in 2017 is a greater focus on machine learning and deep learning in order to automate much of the complexity of predictive analytics, bringing the results into the hands (and minds) of millions of business analysts rather than the much rarer world of data scientists.

Machine learning is also one of the drivers for growing adoption of Apache Spark, the in-memory data processing framework, but there are many others, and it is in part due to the multi-functional capabilities of Spark that we expect to see it overtaking MapReduce in the long term as the default data processing engine for Hadoop. That might not happen in 2017, but the momentum is set to continue in the coming year and Spark is clearly the go-to engine for new Hadoop-based development projects.

Democratizing access to data has been one of the central themes of the rise of self-service analytics and data preparation in recent years, reducing the burden on IT to prepare data sets for business analysts and enabling them to access, prepare and analyze data independently. While there are some advantages to this approach, it has become clear during 2016 that, especially when it comes to data preparation in multi-function data lake environments, there is a balance to be struck between the flexibility provided by self-service approaches and the need for data stewards to retain control over data governance, particularly for privacy, security, regulatory and data quality reasons.

For that reason, we expect to see greater focus in 2017 on what we are calling managed self-service approaches to data preparation. These combine self-service data discovery, cleansing, wrangling, enrichment and matching capabilities for business analysts and data scientists with a managed approach to data governance, cataloging, lineage, security, quality and inventory for data stewards and IT.

The cloud is clearly an ongoing driver for change, and is beginning to have an impact in areas where concerns around security and privacy have previously slowed adoption. One of those is corporate performance management (CPM), including applications for financial planning and analysis (FP&A), financial reporting, consolidation and other performance management tasks. While the potentially sensitive nature of those applications may have discouraged cloud-based adoption in the past – other than by small firms and departments – it seems likely that 2017 could be the year when CPM cloud services enter the mainstream to become a strategic and fundamental part of an enterprise-wide software fabric.

Our final trend this year also deals with the ongoing maturation of an existing technology category – the in-memory data fabric. In-memory data grid/cache technologies have been around for many years, used by enterprises to improve the performance of applications by relieving the load on the underlying database. Many vendors have now improved the functionality of their offerings to the extent that they are increasingly able to claim the potential to use an in-memory data fabric as a replacement for, rather than a complement to, disk-based databases.

Not all companies would truly consider throwing away their disk-based database investments in favor of an in-memory data fabric, but some are, especially as the emergence of new data platforms – such as Hadoop – provides an alternative to disk-based databases for long-term persistence.

451 Research's 2017 Data Platforms & Analytics Trends

Source: 451 Research, 2016

	WINNERS	LOSERS
Machine Learning and Deep Learning Will Enter a New Phase of Strategic Adoption for Predictive Analytics	Vendors that can develop enterprise predictive applications that conform to user needs; companies that take the entire predictive analytics process into consideration from machine-learning testing to deployment	Firms that stick to serving the traditional audience of data scientists and developers; analytics vendors that don't evolve to embrace more advanced analysis use cases
Spark Will Become the Dominant Technology Behind Use Cases for Hadoop	The current group of Hadoop distributors; NoSQL database vendors	Vendors and customers that don't treat Spark as a maturing technology; enterprises that look to reap Spark's benefits but aren't ready to commit to the complexity of operating it
Managed Self-Service Will Balance Agility and Governance Requirements	Self-service analytics and data preparation vendors that balance the requirements of data stewards and IT with the potential benefits of empowering data analysts and data scientists to prepare and analyze their own data sets; data lake management vendors that have invested in the ability to catalog and provide access to data in multifunction environments	Vendors that are only now delivering their 'first wave' of products and services targeted at individual data analysts and data scientists; self-service analytics and data preparation vendors that fail to address the need to deliver more advanced management and governance functionality
Corporate Performance Management in the Cloud Will Truly Come of Age	Organizations that successfully demonstrate the value of performance management to enterprises; vendors that embrace Microsoft Excel provide value-add rather than trying to replace it	Companies that fail to serve use cases beyond bread-and-butter departmental enterprise and mid-market budgeting tasks; organizations that don't meet modern, flexible, collaborative web-based requirements
The Data Grid Will Morph into a Data Platform in Its Own Right	Companies already in the data grid market; end-user companies renewing their database licenses	Relational database companies that do not have a foot in the data grid market; open source projects and proprietary data grid firms that have been less successful at elevating their technologies beyond more straightforward caching

METHODOLOGY

Reports such as this one represent a holistic perspective on key emerging markets in the enterprise IT space. These markets evolve quickly, though, so 451 Research offers additional services that provide critical marketplace updates. These updated reports and perspectives are presented on a daily basis via the company's core intelligence service, 451 Research Market Insight. Forward-looking M&A analysis and perspectives on strategic acquisitions and the liquidity environment for technology companies are also updated regularly via Market Insight, which is backed by the industry-leading 451 Research M&A KnowledgeBase.

Emerging technologies and markets are covered in 451 Research channels including Business Applications; Cloud Transformation; Data Platforms and Analytics; Datacenter Technologies; Development, DevOps and IT Ops; Enterprise Mobility; European Services; Information Security; Internet of Things; Mobile Telecom; Multi-Tenant Datacenters; Networking; Service Providers; Storage; and Systems and Software Infrastructure.

Beyond that, 451 Research has a robust set of quantitative insights covered in products such as Voice of the Enterprise, Voice of the Connected User Landscape, Cloud Price Index, Market Monitor, the M&A KnowledgeBase and the Datacenter KnowledgeBase.

All of these 451 Research services, which are accessible via the web, provide critical and timely analysis specifically focused on the business of enterprise IT innovation.

For more information about 451 Research, please go to: *www.451research.com*.

Table of Contents