



## EXECUTIVE OVERVIEW

**THOUGHT  
LEADERSHIP**

# Total Data Management

**FEB 2017**

**Matt Aslett, Research Director**

This report explores the emergence of continuous data integration, and the role that self-service data preparation, data catalogs and data governance have to play in making the data lake concept a reality.



## 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 over 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.

© 2017 451 Research, LLC and/or its Affiliates. All Rights Reserved. Reproduction and distribution of this publication, in whole or in part, in any form without prior written permission is forbidden. The terms of use regarding distribution, both internally and externally, shall be governed by the terms laid out in your Service Agreement with 451 Research and/or its Affiliates. The information contained herein has been obtained from sources believed to be reliable. 451 Research disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although 451 Research may discuss legal issues related to the information technology business, 451 Research does not provide legal advice or services and their research should not be construed or used as such. 451 Research shall have no liability for errors, omissions or inadequacies in the information contained herein or for interpretations thereof. The reader assumes sole responsibility for the selection of these materials to achieve its intended results. The opinions expressed herein are subject to change without notice.

### NEW YORK

1411 Broadway  
Suite 3200  
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

Paxton House  
(Ground floor)  
30, Artillery Lane  
London, E1 7LS, UK  
P +44 (0)207.426.1050  
F +44 (0)207.657.4510

### BOSTON

75-101 Federal Street  
5th Floor  
Boston, MA 02110  
P 617-598-7200  
F 617-357-7495

## 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

Continuous data integration enables organizations to increase the frequency with which they analyze data in order to deliver greater business agility and benefits, such as accelerated development times and improved customer service.

---

The emergence of the new breed of integration-as-a-service and continuous data integration providers is driven by developers' increasing involvement in data processing and the decreased reliance on data-integration professionals, as well as the increased volume of data generated by SaaS applications.

---

Similarly, the early self-service data preparation vendors tapped into latent business analyst and data analyst demand for self-service data preparation. Established incumbent vendors had failed to identify this demand due to their focus on IT and data management professionals as the primary users of their products.

---

Self-service data preparation can reduce the burden on IT to prepare data for end users, and in doing so reduce the time and overhead for users to discover, integrate, cleanse and enrich data to make it suitable for analysis. This makes it possible for users to expand the scope of their analysis – accessing more data sets and greater volumes of data. It can also expedite the time to value via machine-learning-driven automation and recommendations.

---

The data catalog is a fundamental enabler not just of the management of the data within a data lake, but also of self-service data preparation. By creating an inventory of exactly what data is in the data lake, the data catalog can help analysts discover data sets while data stewards apply data quality and master data management tools to ensure that the data is fit for analysis.

---

# Executive Summary

## INTRODUCTION

In *Total Data Integration*, published in November 2013, we outlined the impact of big data on the data integration sector, including the role that integration and governance technologies had to play in enabling the analysis of structured, semi-structured and unstructured data. In this follow-up report, we examine the changes that have occurred in the data integration and management space since then, including the emergence of the concept of the data lake, as well as the role that self-service data preparation, data catalogs and data governance have to play in making that concept a reality. The report also explores the emerging trend toward ETL-as-a-service and continuous data integration to support enterprise requirements for integrating data from multiple sources, including SaaS and cloud applications, as well as the Internet of Things (IoT).

## METHODOLOGY

This report was generated through research conducted on an ongoing basis from approximately mid-2015 to early 2017. The research included briefings and inquiry sessions with vendors and enterprises, as well as attendance at multiple industry events. In addition, particular effort was applied to identifying the key vendors offering self-service data preparation and data lake management products and services.

Research reports, briefing notes and marketing materials were assessed and re-assessed in order to:

- segment the relevant vendors, products and services into four groups reflecting similar product strategies
- identify the five key areas of functionality (inventory, discovery, preparation, automation and governance) against which these vendors and products could be assessed
- score the vendors and products according to their ability to deliver the five key areas of functionality

Reports such as this one represent a holistic perspective on key emerging markets in the enterprise business applications 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](http://www.451research.com).

# Table of Contents

<b>1. THE CHANGING FACE OF DATA MANAGEMENT</b>	<b>1</b>
<i>Figure 1: The Data-Driven Decision Journey</i> . . . . .	2
<b>2. THE CASE FOR CONTINUOUS DATA INTEGRATION</b>	<b>3</b>
WHAT IS CONTINUOUS DATA INTEGRATION? . . . . .	3
<i>Figure 2: The Continuous Integration Process</i> . . . . .	4
KEY TECHNOLOGIES AND VENDORS . . . . .	4
EVOLVING USER REQUIREMENTS . . . . .	7
<b>3. SELF-SERVICE DATA PREPARATION, DATA GOVERNANCE AND THE DATA LAKE</b>	<b>8</b>
WHAT IS A DATA LAKE? . . . . .	8
DATA GOVERNANCE AND DATA CATALOGS . . . . .	9
DATA CATALOGS AND SELF-SERVICE DATA INTEGRATION . . . . .	10
<i>Figure 3: Self-Service Data Preparation, Data Governance and the Data Lake</i> . . . . .	11
<i>Figure 4: The Virtuous Circle of Inventory, Discovery, Preparation, Automation and Governance</i> . . . . .	11
MOVING TOWARD MANAGED SELF-SERVICE . . . . .	12
KEY TECHNOLOGIES AND VENDORS . . . . .	12
<i>Figure 5: The Self-Service Data Preparation, Data Catalog and Data Lake Management Landscape</i> . . . . .	13
<i>Analytics with Data Preparation Capabilities</i> . . . . .	14
<i>Specialist Self-Service Data Preparation Products</i> . . . . .	15
<i>Combined Data Preparation and Data Lake Management</i> . . . . .	15
<i>Specialist Data Catalog/Data Lake Management Products</i> . . . . .	16
<b>4. CONCLUSION AND RECOMMENDATIONS</b>	<b>17</b>
<b>5. FURTHER READING</b>	<b>18</b>
<b>6. INDEX OF COMPANIES</b>	<b>19</b>