# Obstacles to AI & Analytics Adoption in The Cloud

January 2020

TRIFACTA®

Researchscape

# Contents

# Introduction

Data has increasingly become a critical component of just about every aspect of business and the amount of data is skyrocketing. In fact, <u>90% of the world's data has been created in the last 2 years</u> and it's expected that by <u>2020, 463 exabytes</u> of data will be created every day from wearables, social media networks, communications (business and consumer), transactions and connected devices. While the explosion in the volume -- and more importantly, diversity of data -- is instrumental in supporting the future of artificial intelligence (AI) and accelerates the automation of data analysis, it's also creating the obstacles that enterprises currently face in their adoption of AI.

In order to understand these obstacles, Trifacta launched a benchmark report in conjunction with Researchscape: Obstacles to AI & Analytics Adoption in the Cloud, to examine how data workers in the U.S. across industries are handling the increased move of data to the cloud, the time constraints endured when preparing data for analytics, artificial intelligence and machine learning initiatives, and the impact these obstacles have on the overall success of these projects.

While organizations have made great strides toward implementing AI/ML in their data and analytics processes, one of the greatest challenges that respondents face is data accuracy. Most believe there is great potential to gain efficiencies and improve data-driven decision-making, but as their use cases continue to increase, there is still much room for improvement to remove the obstacles to adoption.
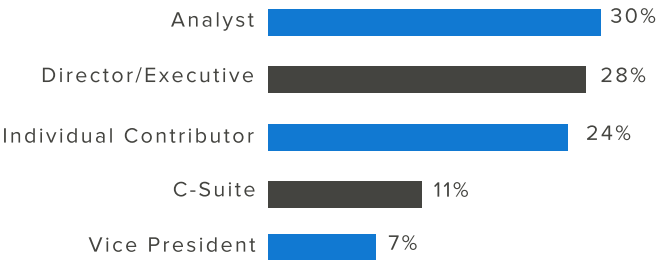
> **"Organizations can no longer rely on legacy compartmentalized data integration to handle the speed, scale, and diversity of today's data."**

# Key Findings

As organizations in every industry modernize their approach to data and analytics, many are moving their data management and analytics to the cloud. Researchscape, on behalf of Trifacta, interviewed over 600 data workers in the U.S. to shed light on the fact that the move to the cloud is an increasingly important one for data storage and management, yet many data workers and leaders face challenges that are impacting not only their bottom lines but also the implementation of AI and machine learning initiatives across their organizations.

When asked about current and expected data usage, 68% of IT professionals reported that all or most of their data is currently stored in the cloud vs. on-premise. About two-thirds of respondents say that all or most of their AI/ML initiatives are currently run in the cloud. Two years from now, nearly nine out of ten of these IT professionals (88%) estimate that all or most of their data will be stored in the cloud.

### What is your title?

| | |
|---|---|
| Analyst | 30% |
| Director/Executive | 28% |
| Individual Contributor | 24% |
| C-Suite | 11% |
| Vice President | 7% |

68% of IT professionals reported that all or most of their data is currently stored in the cloud vs. on-premise.

**88%**
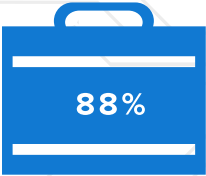
of IT professionals think most or all of their data will be in the cloud.

The survey results also make it clear that poor data quality presents a huge hindrance to AI and analytics adoption in the cloud. While 76% of C-Suite executives have AI and machine learning initiatives included in their company's roadmap, 75% aren't confident in the quality of their data. Poor data quality caused AI/ML projects to take longer (38%), cost more (36%) and fail to achieve the anticipated results (33%). This presents a critical problem, as data is king with decision making. In fact, 71% respondents stated that their company is frequently leveraging data analysis to drive future business decisions.

Implications of data inaccuracy commonly include miscalculating demand (59%) and targeting the wrong prospects (26%). In an attempt to achieve greater accuracy, 46% of respondents spend over 10 hours properly preparing data for analytics and AI/ML initiatives, while some are even spending over 40 hours. In addition, 60% of C-Suite respondents currently use AI to prepare data.

**76%**

of C-Suite executives have AI & ML initiatives in their company's roadmap.

**75%**
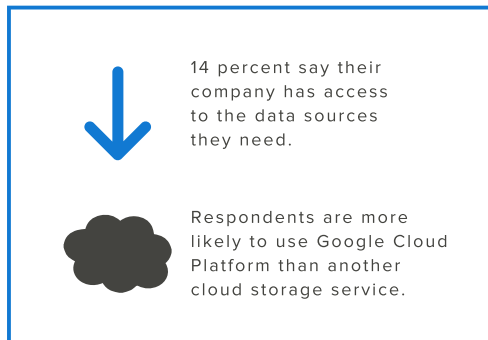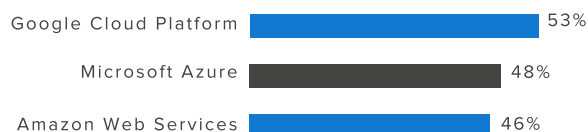
aren't confident in the quality of their data.

**50%**

Nearly 50% (46%) of respondents spend over 10 hours properly preparing data for AI/ML initiatives.

# Rise of AI and ML Push Cloud Adoption

The rise of technologies like AI and ML is propelling a move to scalable, cloud data warehouses and data lake architectures. This not only gives organizations better economics, but also increases their agility, allowing them to focus on driving value from analytics rather than managing infrastructure. IT professionals in organizations using the cloud describe the following benefits of cloud computing: accessibility between environments (51%), increased efficiency (49%), and cost reduction (48%). More than six out of ten respondents (61%) say that AI/ML initiatives are on their company's roadmap.

Google Cloud Platform ████████████████ 53%
Microsoft Azure ███████████████ 48%
Amazon Web Services ██████████████ 46%

14 percent say their company has access to the data sources they need.

Respondents are more likely to use Google Cloud Platform than another cloud storage service.

As cloud platforms continue to improve flexibility and scale, organizations need a modern, self-service approach to exploring & preparing data natively in the cloud to keep up with the business demands for differentiated data. About seven out of ten business professionals surveyed (69%) reported that their organization uses cloud infrastructure for data management. According to the data findings, those who use cloud storage are most likely to use Google Cloud Platform (53%), Microsoft Azure (48%), or Amazon Web Services (46%).

More than seven out of ten respondents (71%) say their company is frequently leveraging data analysis to drive future business decisions. However, only 14% claim to have access to all the data sources they need.

# Data: AI's Best Friend and Biggest Foe

One of the greatest challenges that respondents face is data accuracy. Only 26% reported that their data is completely accurate before preparation and cleaning. Organizations are quickly realizing that AI initiatives are rendered useless, and in some cases detrimental, without clean data to feed their algorithms.
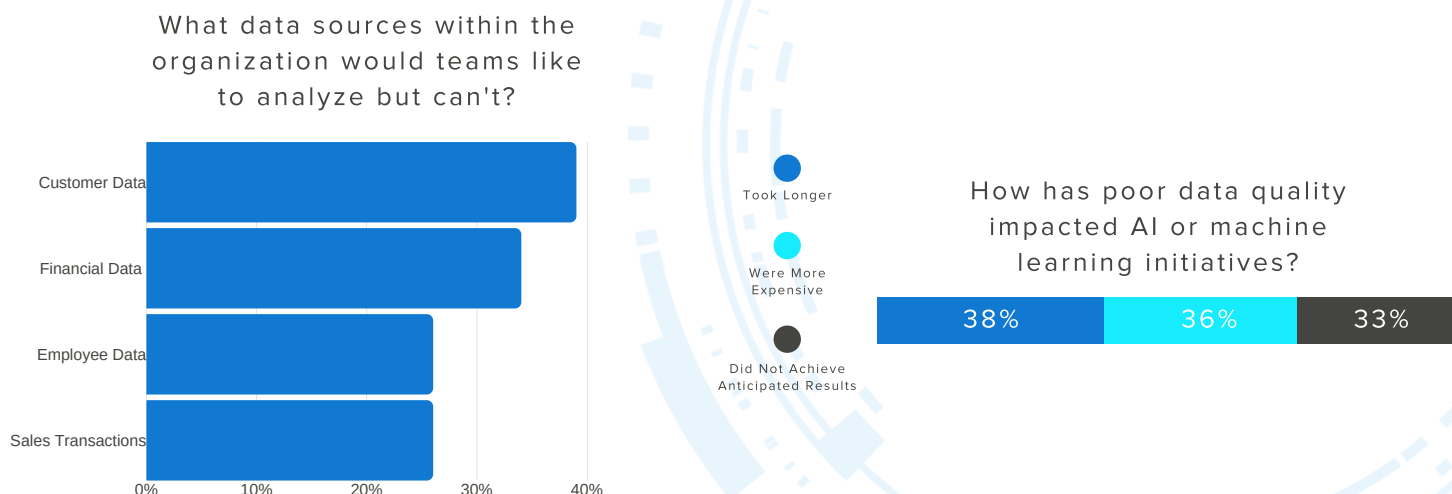
Implications of data inaccuracy commonly include miscalculating demand (59%) and targeting the wrong prospects (26%). Often data accuracy would increase if organizations were able to analyze third-party data from customers, semi-structured data, or data from relational databases. However, common barriers to access include data that exists in different systems (28%) or requires merging from different sources (27%) or needs reformatting (25%). Sought-after data sources include customer data (39%), financial data (34%), employee data (26%), and sales transactions (26%). Furthermore, third-party and secondary data present their own sets of challenges, with about half of respondents citing data blending, data movement, and data cleaning as frequent obstacles.

Only 26% reported that their data is completely accurate.

Poor data quality has had a negative impact on AI/ML initiatives, with more than one-third reporting that projects took longer (38%), were more expensive (36%), or did not achieve anticipated results (33%). Time required to properly prepare data for an AI/ML initiative ranged widely with about one-quarter each saying 1 to 4 hours (28%), 5 to 9 hours (25%), 10 to 19 hours (22%), and 20+ hours (24%). These findings are consistent with Forrester Research's report that up to 80% of an analyst's time is spent on data cleansing and preparation.

### What data sources within the organization would teams like to analyze but can't?

| | |
|---|---|
| Customer Data | |
| Financial Data | |
| Employee Data | |
| Sales Transactions | |

0%   10%   20%   30%   40%

● Took Longer
● Were More Expensive
● Did Not Achieve Anticipated Results

### How has poor data quality impacted AI or machine learning initiatives?

| 38% | 36% | 33% |
|---|---|---|

**Data preparation is cleaning, structuring, and enriching raw data into a desired output for analysis. Data cleansing is the first step in the overall data preparation process and is the process of analyzing, identifying, and correcting messy, raw data.**

# Data Accuracy is the Only Way Forward

Organizations can no longer rely on legacy, compartmentalized data integration to handle the speed, scale, and diversity of today's data. Inadequate data cleansing and data preparation frequently allow inaccuracies to slip through the cracks. This is not the fault of the ETL developer, but a symptom of a much larger problem of manual and partitioned data cleansing and data preparation. According to Harvard Business Review, "Poor data quality is enemy number one to the widespread profitable use of machine learning."

A clean dataset is critical for AI and ML projects, but as sources of data increase, both in the cloud and on-premises, it's challenging for enterprises to combat the problems caused by data inconsistencies and inaccuracy. Innovative data preparation technology can help organizations improve data quality and accuracy for AI/ML initiatives and beyond while also increasing the speed and scale of these efforts. Survey respondents' concerns and priorities for the future speak to how integral these new solutions will become as more organizations rely on data analysis to drive business decisions.

The transformational opportunities provided by the advent of AI and cloud computing will only be available to the extent that organizations can make their data usable. After preparation and cleaning, data accuracy increases to 80% (completely = 29%, very accurate = 51%).  deduplication (21%), data validation (21%), and analyzing relationships between fields (20%) are the most likely steps to improving data accuracy.

Data cleansing can be difficult, but the solution doesn't need to be. Self-service data preparation tools are solving these problems and helping organizations get the most value out of their data with proper data cleansing.

> **"Poor data quality is enemy number one to the widespread profitable use of machine learning."**

Looking ahead, given the implications of data inaccuracy and data quality, organizations would benefit from modern data preparation tools to ensure clean, well-prepared data is always available to support business intelligence, analytics, and AI/ML initiatives across the entire organization.

## Trifacta Cloud Data Preparation

By automating complex data engineering tasks and enabling increased levels of self-service, Trifacta's Cloud Data Preparation platform empowers data professionals of all backgrounds to prepare and clean diverse data for analysis. Through a unique combination of human-computer interaction, machine learning and scalable data processing, Trifacta customers are modernizing their approach to creating and managing their data pipelines.

Unlike other data preparation solutions, Trifacta guides users through the process of exploring and preparing data using intelligent suggestions powered by machine learning. This allows organizations to quickly detect and remediate challenging data quality problems, so they can focus more time and resources on gaining insights and deriving real value from their data. Business units and organizations can spearhead their own initiatives, leveraging new data sources while maintaining the same resources on their team.

## About the Study

**The results in this report are from an online survey that was fielded from August 20 to 30, 2019 by Researchscape International and Trifacta.**

## About Researchscape International

Researchscape International is an agile survey-research consultancy and SaaS firm delivering marketers and agencies PR surveys, omnibus surveys, automated reporting tools and other research-related services. Its surveys are frequently used to drive thought leadership, support content creation and help grow organizations' public profiles. Custom surveys support product launches, crisis communications, customer satisfaction and more.

## About Trifacta

Trifacta is the global leader in data wrangling. Trifacta leverages decades of innovative research in human-computer interaction, scalable data management and machine learning to make the process of preparing data faster and more intuitive. Around the globe, tens of thousands of users at more than 10,000 companies, including leading brands like Deutsche Boerse, Google, Kaiser Permanente, New York Life and PepsiCo, are unlocking the potential of their data with Trifacta's market-leading data preparation solutions. Learn more at trifacta.com.