

**A Yellowbrick  
Data Research  
Paper**



# **Key Trends**

**in Hybrid, Multicloud and  
Distributed Cloud for 2021**

# Contents

**03**

**Executive Summary**

**04**

**Methodology**

**05**

**Key Data Warehousing  
Trends for 2021**

**06**

**A Move Toward  
Analytics Infrastructure  
Investments**

**07**

**Data Warehouse  
Deployment Trends**

**08**

**Biggest Data  
Warehousing and  
Analytics Priorities**

**09**

**Getting More Value  
from Data Lakes**

**10**

**From On-Premises  
to Cloud to Hybrid**

**12**

**Trends Toward  
Distributed Cloud**

**14**

**The Need for a  
Real-Time Analytics**

**15**

**Insights on Future Data  
Warehousing Needs**

**16**

**Summary**

# Executive Summary



**As companies work to adjust their businesses in response to everything that's happened since the Covid-19 pandemic started in 2020, it's not surprising that key drivers of the response are turning out to be data analytics.**

The unprecedented changes (personal, social, and business) have required organizations to re-examine their business processes and existing assumptions rapidly and deeply, from what they sell, to how they sell it, to how they take advantage of future opportunities.

That's where data and analytics come in. To respond effectively to this rapidly-changing business environment, organizations need to be able to confidently make decisions based on data. That makes data, data warehousing, and analytics a key focus area for many companies in 2021 and beyond.

To better understand the immediate impact of all these unprecedented changes on businesses and their IT infrastructures, Yellowbrick Data recently surveyed 500 enterprise IT managers and executives and uncovered their data warehousing and analytics priorities.

The survey revealed some initially surprising results, given all the challenges that organizations are coping with.

For example, it highlighted that even in light of the turbulent business environment, organizations are continuing to place a priority on data analytics and data warehousing. In fact, 76% of respondents noted that they are investing more in 2021 in their analytics platforms.

At the same time, the survey revealed key data warehousing-related trends for 2021, from hybrid and multicloud data warehousing (53%) to real-time analytics (48%) to support for edge analytics (28%). Based on those responses, it's clear that organizations are looking forward.

These areas are potentially huge opportunities for organizations with the right data warehouse platform. For the organizations that are still focused on data warehouse modernization or data lake augmentation, the survey results examined in this paper can help provide a roadmap for their future data warehousing platform capabilities.

Even given all the challenges over the last year or two, the opportunities for using data warehousing and analytics have never been bigger.

With the right solution, the benefits are immediate: faster time to insight, architectural simplification, greater security, increased agility, support for distributed clouds and edge computing, and much more. The results of the survey outlined in this paper can help guide organizations to understand what the future of data warehousing and analytics looks like for 2021 and beyond.

# Methodology

The Yellowbrick ITDM Data Warehousing Survey was conducted in March of 2021. More than 500 IT decision-makers, managers through CEOs, responded. All respondents work full-time for enterprises with 1,000+ employees.



# Key Data Warehousing Trends for 2021

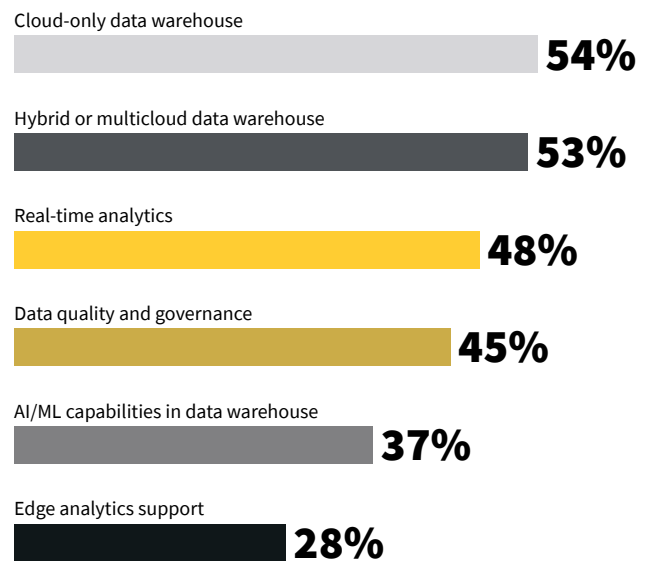
## When asked to select the most important data warehousing related trends for 2021, a few important ones stand out.

The first is expanding deployment flexibility for data warehouses. While 54% of respondents said that conventional cloud data warehouses are a key trend for 2021, a very close 53% stated that hybrid or multicloud data warehouses are an important trend.

A second key trend is real-time analytics (48% response rate). For years, organizations have been limited by analytical processing speed or analysis of historical data. But given all the changes happening, that's no longer good enough. Organizations that want to remain competitive must move to incorporate real-time data sources where they make sense, and they must remove the traditional time lag associated with analytics that can take hours or longer to return results.

While a trend toward data quality was noted by 45% of respondents, a particularly interesting finding was that 28% cited edge analytics support as important. Most organizations are just at the start of considering edge computing deployments, and the fact that 28% are

## What are the most important data warehousing-related trends for your company this year? Select all that apply.



considering the data warehousing implications for edge computing for IoT, mobile, or factory-floor implementations highlights that more companies are thinking about distributed cloud deployments and bringing analytics to the data, rather than the data to the analytics.

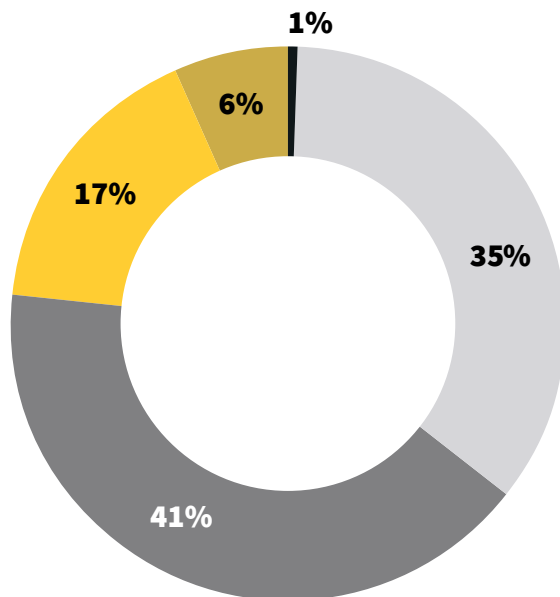
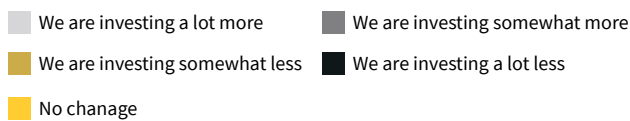
## YELLOWBRICK TAKEAWAY



Key trends for data warehousing in 2021 are basically focused on flexible locations, real time, and increased quality. According to responses, companies are very focused on deploying data warehouses in cloud, multicloud, and hybrid environments. This focus on data warehouse location is expanded to now include edge computing analytics, where analysis can be pushed out to new IoT, mobile, or factory-floor locations. At the same time, real-time analytics is clearly important to a significant number of companies hoping to maintain or grow their competitive position. And as data sources continue to grow, organizations continue to cite the importance of data quality and governance.

# A Move Toward Analytics Infrastructure Investments

In the current climate of uncertainty, our investment in analytics infrastructure such as data platforms, data warehouses, etc. is:



**If you want to understand something, it's said you should follow the money. Thus, when it comes to understanding the role of data, analytics, and data warehousing it's instructive to see what companies are doing with their budgets.**

The results from Yellowbrick's survey of IT executives show overwhelming that organizations see increased value in data warehousing and analytics.

Over 76% of IT executives stated that even in this time of uncertainty, their organization was investing more in analytics infrastructure such as data platforms, data warehouses, and more.

Specifically, over 35% stated they were investing a lot more in analytics infrastructure this year, while another 41% stated they were investing somewhat more. Only 17% noted there was no change.

## YELLOWBRICK TAKEAWAY



If you want to know what's going to be important in the future, follow the money. For 2021, results from the Yellowbrick Data survey indicate the vast majority of organizations (over 76%) are investing more money than before in their analytics infrastructure. That indicates that a clear majority of IT executives feel that even in an uncertain time, investments in data warehouse modernization, real-time analytics, multicloud data warehousing and other analytics infrastructure investments will have a positive impact on their bottom line.

# Data Warehouse Deployment Trends

**For years, data warehouses lived where all other IT systems lived: on premises. That's no longer the case, as results from the Yellowbrick Data survey highlight.**

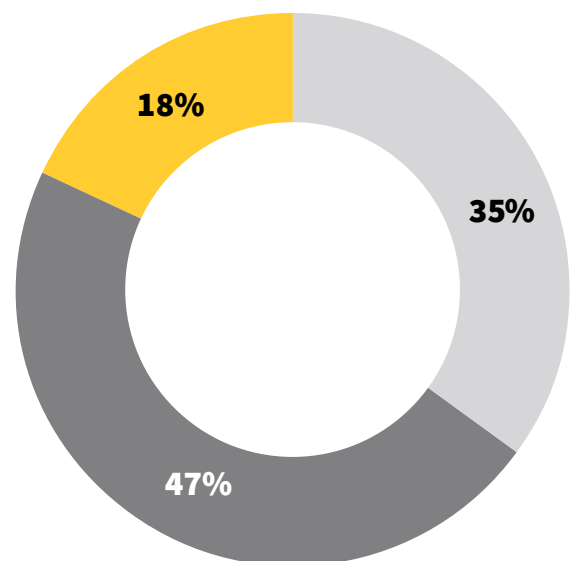
In fact, only 18% of survey respondents stated that all of their data warehouses were on premises, a surprisingly low number given the that ten years ago that number would have been much closer to 100%.

On the other extreme, 47% noted that all of their data warehouses were in the public cloud, with 35% stating they had a mixture of data warehouses on premises and in the public cloud.

For the respondents that had moved (or are moving) their data warehouses to the public cloud, the main reasons cited were to increase the overall data warehouse performance (31%) as well as improve agility for faster innovation (29%).

**Which statement best fits your current data warehouse deployment strategy?**

- We have data warehouses on-premises and in the public cloud
- All of our data warehouses are in the public cloud
- All of our data warehouses are on-premises



## YELLOWBRICK TAKEAWAY



Data isn't useful in itself until it's turned into knowledge and acted upon with decisions. Deployment trend results from the Yellowbrick Data survey show this in action—organizations are migrating their data warehouses to the public cloud to increase performance (to gain insights faster) and improve agility (for faster decisions and innovation). Another key take-away from the results highlights the importance of having an analytics strategy that can encompass a mixture of deployment locations.

# Biggest Data Warehousing and Analytics Priorities

## The Yellowbrick Data survey shows that organizations are prioritizing and investing in their analytics infrastructure in 2021.

They're also continuing the deploy data warehouses in a mixture of environments, from private data centers, to public clouds, to hybrid clouds.

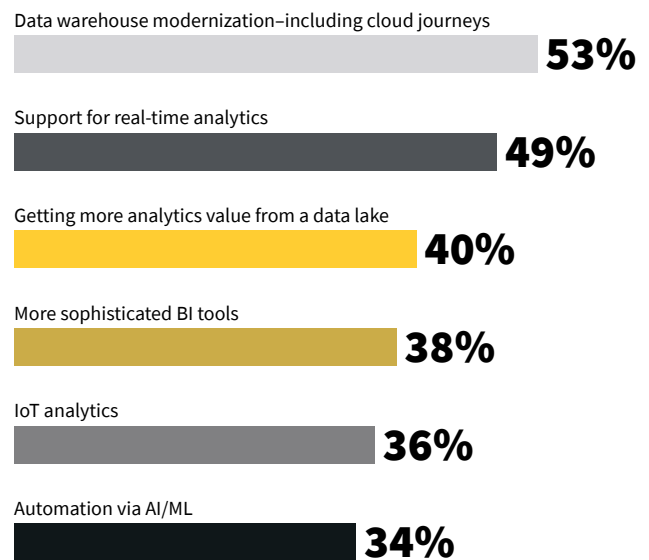
But what are the priorities driving these investments and decisions? What specifically are organizations hoping to gain from such investments and architectural changes?

Two main things, it turns out. When asked what summarized the primary improvement that was required for their company's data warehouse in 2021, two important reasons took the lead: to easily scale to support more data and users (31%), and to support processing in the cloud, on-premises, and at the edge in an integrated way. (28%)

Other less important (but still notable) reasons included reduced query latency for faster insights (11%), cheaper and easier to manage (13%), and ingest and process real-time data with ease (17%).

These answers make sense, given the increased importance of data and analytics in today's business environment, and the growing role of cloud-based platforms in most organization's business strategies.

## What are your highest analytics-related IT budget priorities in 2021?



A follow-up question focused on organizations' highest analytics-related IT budget priorities for 2021. Here, IT executives called out data warehouse modernization (53%) as their highest priority, followed by support for real-time analytics (49%). The other most interesting priorities were getting more analytics value from data lakes (40%) and IoT analytics (36%). We'll dive into getting more analytics value from data lakes in the next section, since it's important in its own right.

## YELLOWBRICK TAKEAWAY

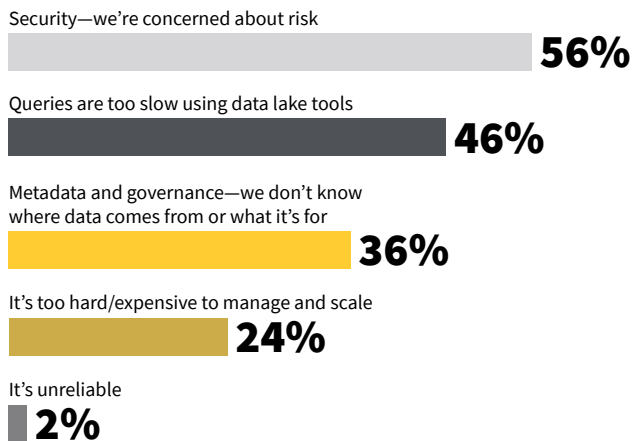


The Yellowbrick Data survey revealed that IT executives are spending money on analytics infrastructure in 2021 for very specific reasons—they need to scale systems to support more data and users, and they want increased (and integrated) deployment flexibility that supports everything from on-premises to cloud to the edge. Since many organizations already have significant data warehouse investments, this basically translates into money for data warehouse modernization, which is why it was noted as the highest IT budget priority. And the need to get more analytics value from data lakes leads directly to the next set of survey responses.

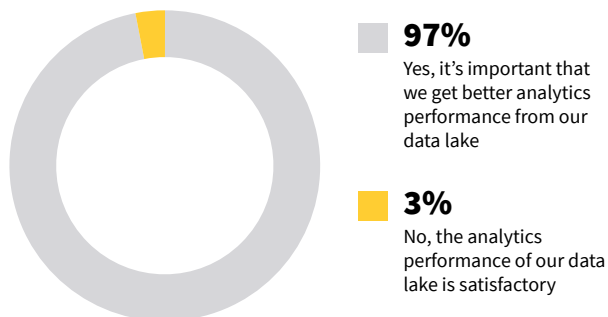


# Getting More Value from Data Lakes

## What are your biggest data lake pain points today?



## Are you actively working to improve analytics performance?



**When you are talking about data lakes, there's probably no need to agree to disagree. Almost no company is getting all the value out of data lakes that they had originally hoped to.**

While data lakes can be a great solution for limited situations, they haven't delivered the low-cost, easy-to-use analytics that most companies had expected. In fact, many of the real-world problems facing companies that are trying to leverage data lakes were highlighted in the survey.

IT executives cited a range of data lake pain points including: security (56%), queries that are too slow (46%), metadata and governance issues (36%), and just too hard to manage and scale (24%).

The end result of all these complications is that the vast majority (97%) of companies say that it's important for them to get better analytics performance from their data lakes.

## YELLOWBRICK TAKEAWAY



The survey revealed that for most organizations, data lakes aren't the answer to the data analytics problem. In fact, data lakes have a range of inherent problems that organizations are actively trying to solve as they plan their data warehouse modernization efforts. The move to cope with data lake problems leads directly into the important trend of distributed cloud, which will be explored more in subsequent survey responses.

# From On-Premises to Cloud to Hybrid

## Where should data warehouse solutions reside? In data centers/private clouds, public clouds, or in both places (hybrid cloud)? Does it depend on whether you have one data warehouse or multiple ones?

This is an especially important question considering how executives responded when asked how distributed or centralized their enterprise data warehouse is. Only 37% answered that they have a single, centralized data warehouse. The same amount (37%) noted they have multiple data warehouses and data marts that are not connected. And 26% have multiple data warehouses and data marts that are connected. That's a total of 63% that have multiple data warehouses or data marts in some form.

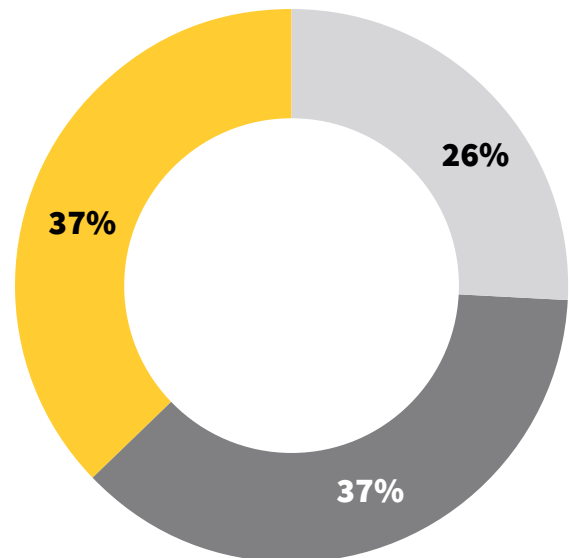
As noted earlier in the survey, many organizations have some, or all, their data warehouses in the cloud. For the ones that do, two key drivers for the move to the cloud were: leveraging benefits such as elastic scalability and high-availability (63%) and leveraging benefits such as ease of management (59%).

On the flip side, organizations that are not (or cannot) use cloud data warehouses are abstaining for a variety of reasons, including: security (57%), unpredictable costs (38%), performance concerns (37%), and government regulations related to cloud data (32%).

Stepping back, though, the vast majority of respondents to the Yellowbrick Data survey noted that the strategic importance of taking a hybrid cloud approach (where the data warehouse run both on the public cloud and on-premises) to data warehousing going forward. In fact, 86% stated that hybrid cloud data warehouse capabilities were critical for their business!

## How “distributed” or “centralized” is your enterprise data warehouse?

- We have multiple warehouses and data marts that are connected
- We have multiple data warehouses and data marts that are not connected
- We have a single, centralized data warehouse



**If your business must move from an on-premise data warehouse to a cloud-based data warehouse, which of the following best describes the reason?**

We want to leverage benefits of the cloud such as elastic scalability and high availability **63%**

We want to leverage benefits of cloud computing such as ease of management **59%**

Our company has issued a cloud-only mandate for data systems **37%**

**If your organization cannot use a cloud-based data warehouse, which of the following best describes the reason?**

We are concerned about security of data in the cloud **57%**

The costs of cloud-based data warehouses **38%**

The performance of cloud-based data warehouses does not meet our business requirements **37%**

We are prevented by government regulations from storing or moving data to the cloud **32%**

Other (please specify) **1%**

## YELLOWBRICK TAKEAWAY



These days, organizations have numerous options when it comes to data warehouse deployments. But one thing is clear from the Yellowbrick Data survey—organizations want a hybrid cloud data warehouse that can seamlessly run in private data centers and/or public clouds. Such solutions can overcome the limitations cited to conventional cloud data warehouses (such as regulatory requirements or security concerns) while also delivering cloud benefits such as scalability, high-availability, and ease-of-management.

# Trends Toward Distributed Cloud

**Today's organizations need to use, incorporate, manage, and analyze data and data sources that spread across a wide range of distributed locations.**

Some data is local. Other data is in a private cloud. More data still is in public cloud sources. And real-time data might exist out at the edge. As they look to remain competitive, organizations must have a data warehousing and analytics strategy that can incorporate all those scenarios, effectively and efficiently.

That's where distributed clouds come in. A distributed cloud is a mesh of interconnected resources (private data centers, public clouds, and edge networks) that are managed in a consistent, unified way.

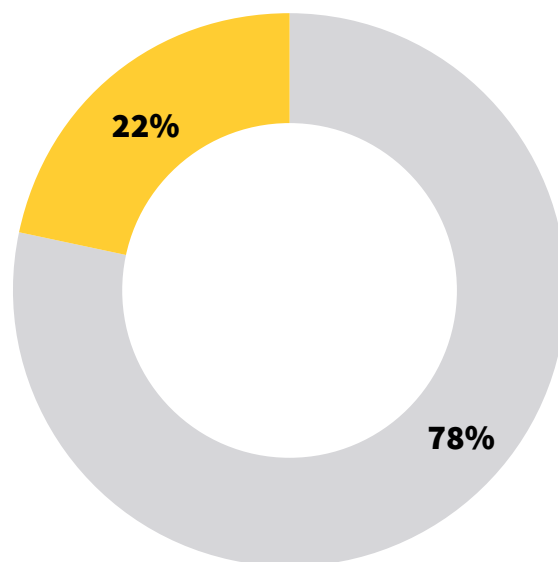
The Yellowbrick Data survey results highlight the importance of distributed clouds. Over 78% of the respondents stated that their organizations leverage distributed clouds. When asked why, the top three responses were:

- Business locations that included remote or disconnected environments and process large amounts of data (58%)
- Need to comply with industry- or country-specific data privacy regulations (52%)
- Need to incorporate IoT or AI applications that run on the edge and must process data in real time.

For businesses that leverage distributed clouds, analytics were an important an important consideration. 48% of those responding cited a need for real-time analytics for large amounts of data in an edge environment in the next six months, while another 39% cited a need within the next 12-18 months. 11% cited a longer timeframe.

Does your business leverage "distributed clouds?"

Yes No



**If your business leverages distributed clouds, which best describes their use?**

Our business includes remote or periodically disconnected locations (such as factories or mobile environments) that generate and process large amounts of data  
**57%**

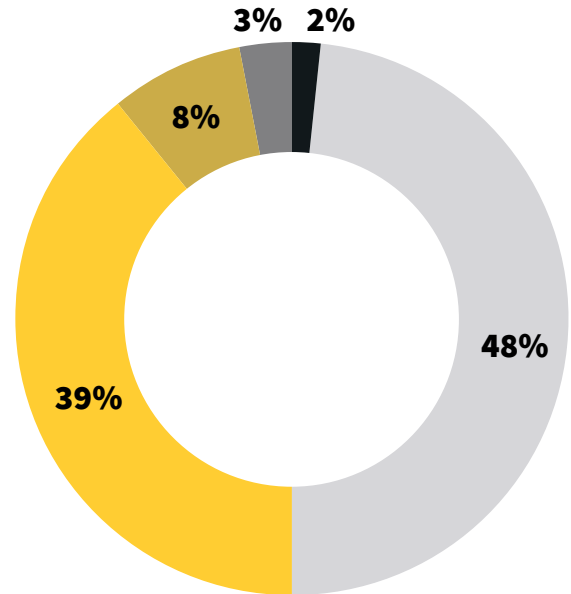
Our business must comply with industry or county-specific data privacy regulations  
**51%**

Our business incorporates IoT or AI applications that run “on the edge” and must process data in real-time  
**50%**

Other (please specify)  
**1%**

**If your business leverages distributed clouds, which best describes your relevant future analytics requirements?**

- We will require real-time analytics processing for large amounts of data in an edge environment in the next 6 months
- We will require real-time analytics processing for large amounts of data in an edge environment in the next 12 to 18 months
- We will require real-time analytics processing for large amounts of data in an edge environment in the next 18 to 36 months
- It will be at least 5 years before we will require real-time analytics processing for large amounts of data in an edge environment
- We will not require real-time analytics processing for large amounts of data in an edge environment in the foreseeable future



**YELLOWBRICK TAKEAWAY**



A surprising percentage (78%) of survey respondents stated that they’re already leveraging distributed cloud environments for a variety of reasons such as regulations, disconnected environments, or edge computing requirements. On top of that, 87% of those businesses identified a strong need for real-time analytics for large amounts of data within their distributed cloud environments within an eighteen-month timeframe.

# The Need for Real-Time Analytics

## One of the big growth areas in IT and business over the past decade has been the explosion of real-time data.

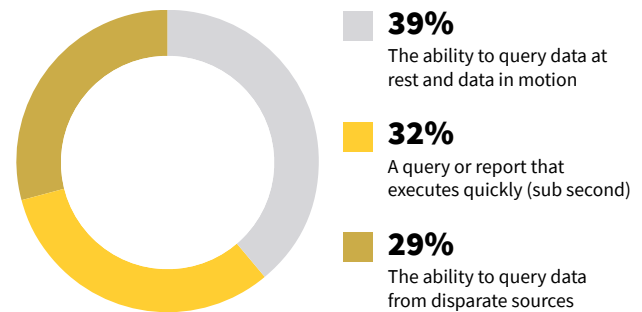
At this point, there's almost no end of real-time data feeds or sources. For example, real-time data feeds can be anything from retail transactions, geo-location data, weather events, manufacturing equipment, fraud analysis, healthcare data, or just about anything else.

But in addition to real-time data, organizations need to consider real-time analytics. To clarify what organizations are looking for in terms of real-time analytics, the Yellowbrick Data survey asked IT executives to define the term. The results were balanced, but split. The respondents almost equally saw real-time analytics as one of three options: the ability to query data at rest and data in motion (39%), a query or report that executes quickly (32%), the ability to query data from disparate sources (29%).

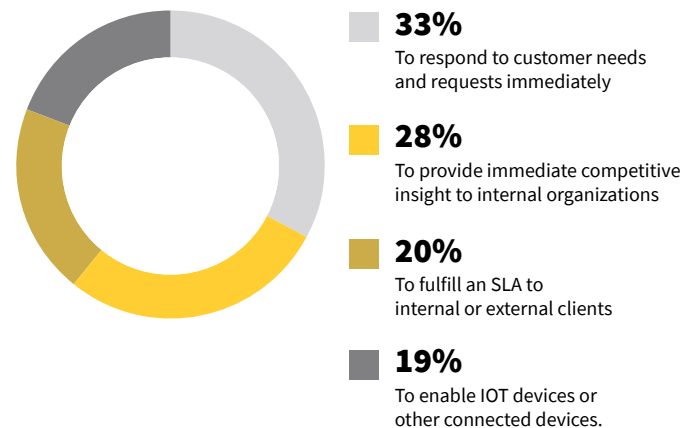
All three choices are clearly valid instantiations of real time analytics, so it's important to understand what your organization's requirements for real-time analytics are, and the right type of analytics and data warehouse infrastructure that can support them.

In addition, respondents were asked a follow-up question on the primary use of real-time analytics within their organization. The two main responses both had to do with fast responses and insights: the most popular use of real-time analytics is to respond to customer needs and requests immediately (33%) and to provide immediate competitive insight to internal organizations (28%).

## What does "real-time analytics" mean to you?



## If "real-time analytics" is important to your business, how would you describe the primary use?



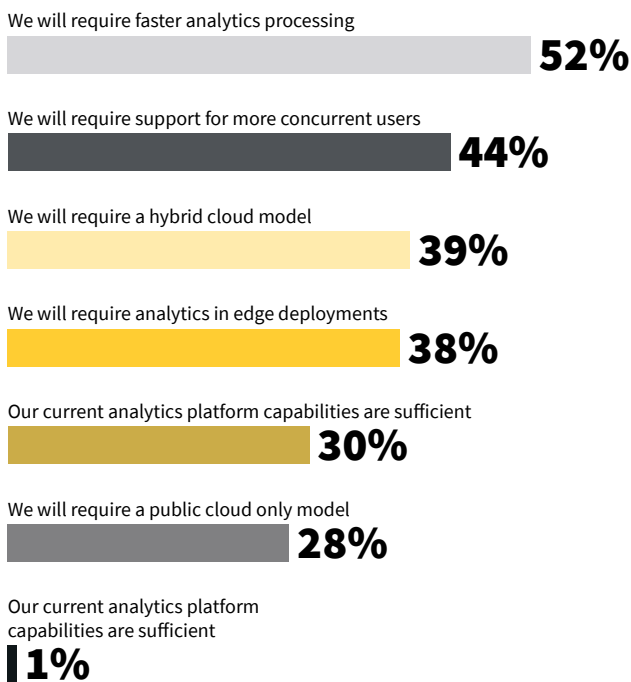
## YELLOWBRICK TAKEAWAY



As real-time data has proliferated over the past decade, the need for real-time analytics has increased. Organizations need to be able to analyze not just real-time data, but they also need data warehousing and an analytics infrastructure that can meet the ever-tightening time requirements for reports, queries, and insights.

# Insights on Future Data Warehousing Needs

## How would you describe the future roadmap for your DTA warehouse capabilities?



While we don't know exactly what the future will hold, based on the responses from the Yellowbrick Data survey, we do know a few things about what the future of data warehousing looks like.

Specifically, organizations were asked to describe the future roadmap for their data warehouse capabilities. And one answer stood out among all others: speed.

52% of the survey respondents, the most for any answer, identified the need for faster analytics- processing as a key item on their data warehousing roadmap.

Three additional requirements were relatively close in response weight: support for more concurrent users (44%), hybrid cloud support (39%), and analytics for edge deployments (38%).

## YELLOWBRICK TAKEAWAY



Given the explosion of data that organizations have to deal with, along with rapidly changing business conditions and the need for greater business insights, it's no surprise that faster analytics would be the core feature for data warehousing roadmaps. Organizations that invest in platforms that provide superior speed and performance are able to run deeper analyses, identify important insights faster, and enable broader use of analytics across the organization. A second important take-away from the responses to this question is the fact that analytics support for edge deployments rated in the top three future data warehousing needs. It shows that organizations are beginning to realize that their future data warehousing platforms need to be able to support distributed cloud environments and much more agile and flexible analytics capabilities can even be pushed out to the edge.

# Summary



**Regardless of your company size, data warehousing and analytics will be a key enabler for future growth and profitability. Yet most organizations are struggling to define a path that will provide the optimal analytics platform for the future.**

To help organizations navigate these areas, Yellowbrick Data created a custom survey of 500 IT decision makers to identify the key trends in hybrid, multicloud, and distributed cloud for 2021.

The results from the survey explored previously in this survey analysis highlight the fact that even though the pandemic has upended business plans across the board, the majority of organizations are continuing to increase investments in data warehousing and analytics to address current and future needs.

The survey also showed that IT leaders want high performance, agility, and security from their data warehouses as they improve hybrid and multicloud environments and move towards new distributed cloud platforms.

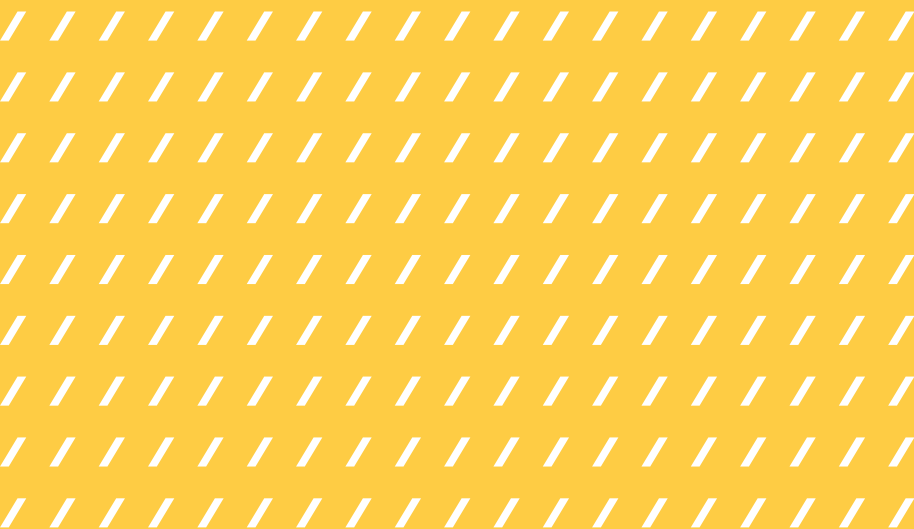
In addition, the majority of organizations (53%) are choosing hybrid or multicloud data warehousing as one of the most important data warehousing trends. Nearly half (48%) said real-time analytics was one of the most important trends, and a surprising number (28%) cited edge analytics support, hinting at a fast-approaching future where new data warehouses and distributed cloud bring the analytics to the data, rather than the data to the analytics.

If your data analytics capabilities or data warehouse isn't keeping up with your business needs or can't address these types of future requirements, it's time to consider modernizing it.

Yellowbrick offers a state-of-the-art data warehouse that can be deployed on any IT infrastructure across private, public, and hybrid clouds (and eventually at the edge in full distributed cloud modes), with all deployments managed via a simple, unified control plane.



To learn more about Yellowbrick Data,  
call us at 877.492.3282 or visit  
[yellowbrick.com](http://yellowbrick.com) to book a demo today.



**Learn more**