THE CLOUD DATA BRIEF
Big Data Transitions to the Cloud
Most businesses strive to capture and analyze their data. But how and where does this happen?

Solutions were once limited to databases deployed exclusively on-premise. But today, businesses must navigate a wide variety of technologies and deployment types. From individual database and Hadoop products to cloud, on-premise, and hybrid deployments, the landscape is diverse, and changing fast.

The Cloud Data Brief is an ongoing project aimed at quantifying the evolution of this landscape. It explores connections to data sources used globally by customers of Tableau Online. Understanding usage trends in these connections gives us a glimpse into how and where businesses are analyzing their data today.

Our Methodology:

The Cloud Data Brief analyzes a sample of over one million anonymized data source connections published to Tableau Online by more than 4,000 customers. As a cloud-hosted tool, usage patterns may represent a bias toward cloud overall.

That said, Tableau Online customers span nearly every industry and vary from small businesses to enterprise corporations. The data in this report includes them all, and we believe it serves as an accurate indicator of broad trends in the market.
Key Findings:

Data storage is expanding beyond our traditional concepts of databases

- Tableau Online customers employ a wide variety of data sources for their analytics. There are over 40 types of data sources, ranging from proprietary databases to open-source Hadoop tools, used by customers. Two-thirds of these data source types are more frequently deployed in the cloud than on-premise.

Data’s center of gravity is moving to the cloud

- Connections to cloud-hosted data have increased by 28% in the past 15 months. In January 2015, data source connections in Tableau Online were distributed somewhat equally between cloud and on-premise deployment. In the first quarter of 2016, that split had swung to 70-30.

- Cloud-native data sources like Amazon Redshift and Google BigQuery are gaining market share. At the beginning of 2014, they represented just 12% of all connections in Tableau Online. By the first quarter of 2016, they had grown to 28% of all connections.
Hybrid data technologies remain critical to business operations

- Hybrid data sources—those that can be deployed on-premise or in the cloud—remain, by far, the most common type of connection used by Tableau Online customers. **Over the past two years, hybrid sources have never dropped below 60% of all connections.**

- For hybrid data, the center of gravity is also edging toward the cloud. Fifteen months ago, Tableau Online customers deployed hybrid data sources evenly across cloud and on-premise environments. **By the first quarter of 2016, connections to hybrid data sources deployed in the cloud had increased to 60%.**

**Want more?**

Read on for a visual analysis of the data, and visit the Cloud Data Brief site for the next release analyzing real-world usage across cloud platforms like Amazon Web Services, Microsoft Azure, and Google Cloud Platform.
The Data Landscape Is Diverse

Tableau Online users employ over 40 types of data sources.

With file-based sources like Excel and business-specific web applications like Salesforce and Google Analytics excluded, there are 32 distinct types of databases and Hadoop ecosystems represented. This diversity is indicative of the wide and varied landscape of data management tools available today.
The Data Landscape Is Diverse

In the future, this landscape will only become more crowded. Gartner reports operational database management systems are experiencing “a dynamic resurgence, with new entrants challenging established leaders.” Gartner also found disruption is accelerating innovation in warehousing tools designed explicitly for data analysis.

User preference for cloud versus on-premise deployment varies widely across data source types.

Two-thirds of the data sources shown above are more often deployed in the cloud than on-premise. But while some data warehouses like Amazon Redshift and Google BigQuery are cloud-exclusive, others like SQL Server and MySQL are deployed across cloud and on-premise environments.

Where are deployment preferences headed overall?
Cloud Data is Trending

Connections to cloud deployments are growing. Data sources deployed in the cloud accounted for 55% of all connections published to Tableau Online in January 2015. In this first quarter of 2016, they made up 70% of all connections. As a percent of total, this growth equated to a 28% increase in just 15 months.
Cloud Data is Trending

What is driving growth of connections to cloud data sources?

The cloud undoubtedly helps organizations reduce costs and save time in provisioning and maintaining infrastructure. But three additional trends are further driving usage of cloud data sources:

• New companies are often “born in the cloud.” These organizations, which deploy their data infrastructure in the cloud from day one, account for a small but growing portion of the market.

• New sources of data necessitate highly scalable environments. The foremost examples is the Internet of Things (IoT), where machine-generated data continually flows from devices and sensors. Organizations often use scalable cloud infrastructure to capture and store this data. Research firm Enterprise Management Associates estimates that IoT records now account for 33% of data in cloud environments. Additionally, these scenarios are becoming commonplace in enterprise settings where IoT data is used to inform and improve business systems.

• Devices like phones and tablets are increasingly used to accomplish business tasks. Dresner Advisory Service’s Wisdom of Crowds 2015 market survey found nearly 70% of respondents named mobile device support as “important, very important, or critical” to business intelligence initiatives. Cloud solutions that make data accessible anytime and anywhere provide the foundation for mobile access of business data.

But while some databases are exclusive to the cloud, other solutions can be deployed locally or in a cloud environment. Which option do businesses choose?
Cloud-Native Grows, But Hybrid Still Dominates

Connections to cloud-native data sources have increased by 125% over the past two years.

At the beginning of 2014, data sources only available in the cloud represented just 12% of all connections in Tableau Online. By March of 2016 they accounted for 27%. This growth is driven by Tableau Online customer’s use of cloud data warehouses like Amazon Redshift and Google BigQuery.

This graph examines connections from January 2014 through March 2016.
Cloud-Exclusive Grows, But Hybrid Still Dominates

But the most common connection type remains “hybrid” data sources.

Hybrid data sources, which can be deployed either on-premise or in the cloud, accounted for 73% of all connections at the beginning of 2014. In March of 2016 they represented 62%. Despite this slight decline, the hybrid category still represents the vast majority of all data source connections published by Tableau Online users.

Why are hybrid data sources so prevalent? Many organizations build analytics based on a mixture of data sources, not all of which can be deployed in the cloud. Security requirements for certain types of financial and healthcare data, for example, often demand on-premise storage. Other times, moving operations to the cloud is an incremental process achieved over months or years. Gartner predicts competition in 2016 will focus on the delivery of these hybrid offerings, which will become the norm by 2018.

Within the hybrid category, which individual database technologies do business users choose to deploy, and where?
The center of gravity for hybrid data sources has shifted to the cloud.

In January 2015, Tableau Online customers connected to hybrid data sources evenly across cloud and on-premise environments.* By the first quarter of 2016, connections to cloud-deployed hybrid data sources had increased, but only slightly, to 60%.

* The Cloud Data Brief uses the anonymous IP addresses of hybrid data source connections in Tableau Online to determine whether they were deployed on-premise or in the cloud. There is a small margin of error associated with this approach, but it correctly identifies the majority of deployment types.
Gartner predicts enterprise resource planning (ERP) environments, which incorporate both cloud and on-premise software deployments, will be commonplace by 2018. Like the prominence of hybrid data overall, the near-even split of hybrid data source deployment across cloud and on-premise indicates that the market, while clearly trending towards cloud, is still very much in transition.

The four most common hybrid data sources used by Tableau Online customers are Microsoft SQL Server, MySQL, PostgreSQL, and Oracle.
Of these four types, SQL Server and Oracle are more likely to be deployed on-premise.

Locally-deployed SQL Server is, on average, twice as common as SQL Server in the cloud. Connections to on-premise Oracle data sources are four times more likely than those in the cloud.

SQL Server and Oracle are both common technologies of the pre-cloud era. They have long been deployed on-premise, and a bias for on-premise deployment is not surprising. However, as Microsoft’s Azure cloud platform grows—Rightscale’s *State of the Cloud* found use of Azure doubled from 2014 to 2015—we may soon see cloud deployments of the company’s SQL Server become more prevalent.

In comparison, connections to cloud-deployed MySQL and PostgreSQL are far more common than on-premise deployments.

Cloud MySQL is, on average, nearly six times more likely than on-premise MySQL. Cloud-deployed PostgreSQL is five times more common than local deployments of PostgreSQL.

As open-source technologies, the popularity of both MySQL and PostgreSQL in the cloud is likely due to their compatibility with all major cloud platforms (including Amazon Web Services, Microsoft Azure, and Google Cloud Platform). Tableau Online customers’ preference for cloud MySQL and PostgreSQL also underscores larger industry trends toward use of open-source database technologies in the cloud.
From deployment options to specific database products, businesses wanting to capture and draw insight from their data are presented with a multitude of choices. On-premise, cloud, or hybrid? Redshift or SQL Server? Selecting the best solution can be a challenge.

But despite the multitude of choices available, a recent survey by Enterprise Management Associates (EMA) found businesses are less interested in the attributes of individual technologies. Instead, they are focused on the time-to-value provided by a given solution. Indeed, the same survey found 40% of organizations believe their data management systems help establish a competitive advantage.
The Data Landscape Of Tomorrow

Cloud-hosted software is increasingly seen as the best way to create this competitive advantage.

To build a successful data strategy, organizations need to understand and account for shifts happening in the world of big data. Industry analyst like Gartner suggest the majority of IT spending will soon be directed at “modernizing, expanding, or substituting long-standing business applications with cloud-based applications.” The growing popularity of cloud-deployed data sources among Tableau Online users indicates the cloud is gaining momentum in the world of big data as well.

Three particular trends are shaping the future of data used for analysis and visualization:

1. **Data’s center of gravity is moving to the cloud.**

   *Data gravity* indicates the pull of data on services and applications. If your data lives in the cloud, you’ll likely want your data tools – from processing to analytics – running in the cloud as well. Data’s center of gravity is now fixated on the cloud, and that focus will only grow larger in the future. Organizations building data ecosystems should concentrate their efforts on cloud workflows to ensure their systems are ready for this change in data gravity.
2. In the move to cloud, hybrid data technologies are critical to business operations.

When not all your data can be moved to the cloud, or you want to make the move incrementally, hybrid data options give you the flexibility to bridge that gap between your cloud-hosted and on-premise environments. Gartner recently predicted these hybrid offerings will become the norm by 2018. For businesses transitioning to the cloud, hybrid is already the norm.

3. Data storage is quickly expanding beyond traditional concepts of databases and warehouse.

Data is now flowing from everywhere and everything. As a result, the storage landscape is expanding to meet the requirements of new and variable streams of data. Cloud-hosted data tools in particular are driving expansion with “as a service” products like Snowflake, as well as trusted services like Amazon’s Relational Database Service. In the future, the landscape will only become more crowded. To capitalize on the breakneck speed of innovation, businesses are building data workflows that focus on flexibility and choice above all else.

Stay tuned for the next release of the Cloud Data Brief, and visit our website for more information on how Tableau works with the databases analyzed in this survey.
Tableau Software helps people see and understand data stored on-premise, in the cloud, or both.
To learn more about how the world of data is shifting, and what it means for your organization,
read our whitepaper:

→ 3 Shifts in the Modern Data Environment, and What it Means for IT Leaders