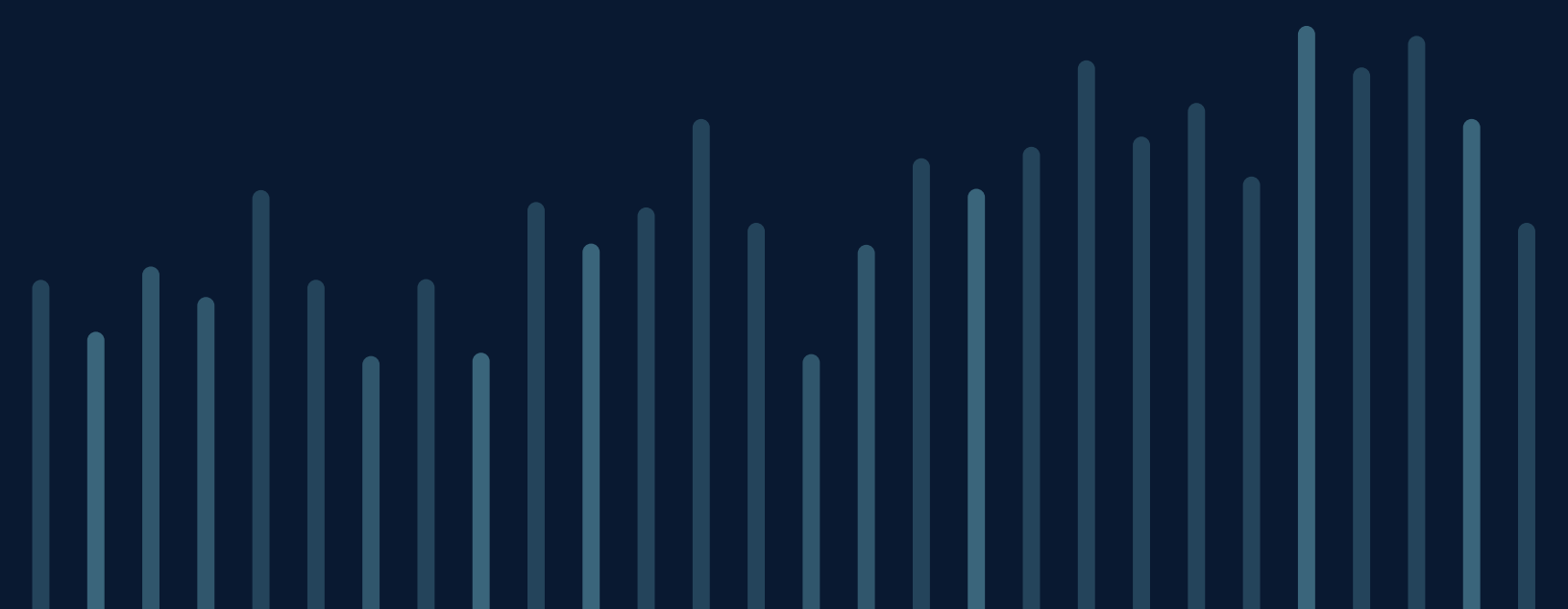




# Tableau Online Scalability

Overview and Proof Points

Ivo Salmre, Director of Product Management, Tableau Online



# Contents

- Introduction ..... 3
- Tableau Online worldwide architecture..... 3
  - Multi-pod, multi-geography architecture ..... 3
  - Site migrations ..... 4
  - Architecture inside a pod ..... 5
  - Security ..... 6
  - Backups ..... 6
  - Resource governance ..... 6
  - Supporting 24/7 mission-critical sites ..... 7
  - Service Level Agreement (SLA) ..... 7
  - Data Management and Resource Blocks ..... 8
- Proof points: Worldwide customer usage ..... 8
- Case study: Tableau’s alpo-dev site on Tableau Online ..... 9
  - How does alpo-dev resemble typical customer enterprise sites? ..... 10
  - How is alpo-dev different from typical customer Online usage? ..... 10
  - Site configuration ..... 10
  - Authentication, user, and group provisioning ..... 10
  - Configuring and provisioning Tableau Bridge ..... 11
  - Monitoring and troubleshooting the alpo-dev site ..... 12
  - Alpo-dev site statistics ..... 12
- Appendix — Q&A ..... 13
- About Tableau & additional resources ..... 15

# Introduction

Part of our mission to help people see and understand data means ensuring our customers have confidence in the scalability and availability of our SaaS analytics platform, Tableau Online. When everyone in your company depends on data and analytics, you can't afford to let them down. And we believe it's important you understand how our hosted service leverages enterprise-grade cloud technology to scale as fluidly as your business grows.

This document describes Tableau Online's high-level architecture and explains how the architecture scales to serve large sites with thousands of geographically distributed users. You will learn how one of Tableau's largest internal sites is configured and performs at scale in Tableau Online.

## Tableau Online worldwide architecture

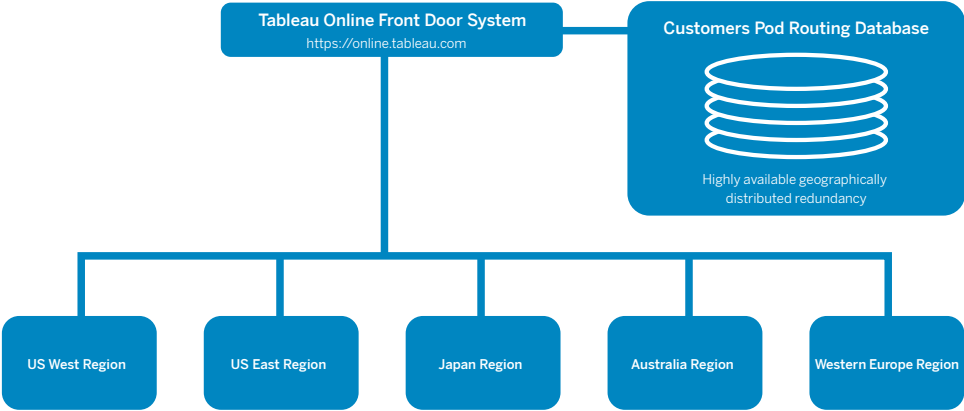
### Multi-pod, multi-geography architecture

Tableau Online has a pod-based architecture hosted on Amazon Web Services (AWS) that supports geographically distant pods. To simplify sign-in, these pods are unified by a common front door system at [online.tableau.com](https://online.tableau.com). All infrastructure is geographically redundant. Tableau Online has multiple customer pods located across the globe in the United States, Western Europe, Japan, and Australia. The Japan and Australia pods were launched in 2020 to better serve customers in these geographic regions.

As is common for pod-based cloud services, each customer's site is homed on a specific geo-located pod. When signing in, a customer is routed to the pod where their site is hosted. Figure 1 below illustrates the architecture.

The customer and user management system, backed by a Customer Pod Routing Database, maps sites

and users to pods. This system also facilitates single organizational identity management by routing to the customer’s single sign-on (SSO) systems (SAML/SCIM/OpenID), such as Okta, Microsoft Azure Active Directory, Ping Identity, or OneLogin.



**Figure 1** Multi-pod, multi-geography architecture

Customers choose their initial geographic pod when creating their site. Tableau can expand the set of pods both within each region as customer growth dictates as well as expand the infrastructure within each pod to service expanded usage.

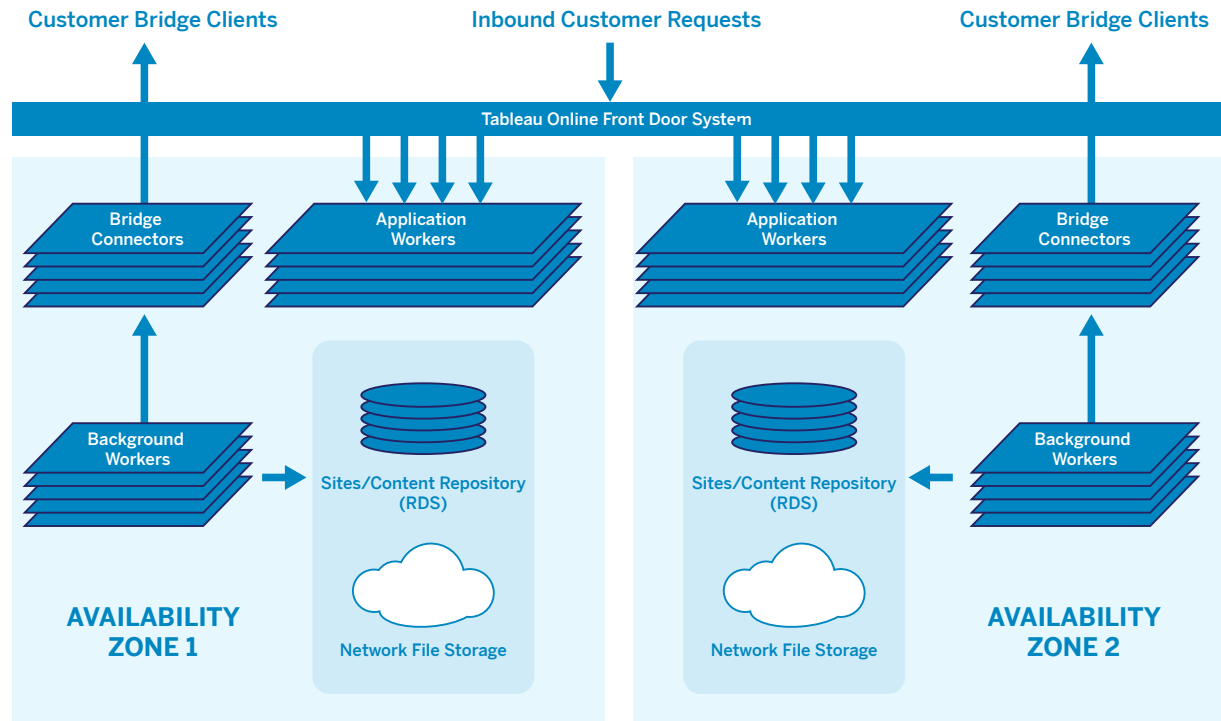
**Site migrations**

Customers occasionally choose to migrate to a new site within the same region or move to a different geographic region. This usually occurs when two organizations have merged, when an organization splits, or when organizations choose to move to a new region based on where their users or data reside.

Tableau offers a self-service process to migrate content (workbooks, data sources, flows) between sites using the Tableau Content Migration Tool. The Content Migration Tool can also be used to move content between on-premises Tableau Servers and Tableau Online; an important capability for customers moving their infrastructure into the cloud. The tools and self-service process are available to customers for the duration of their migration between sites or between Tableau Server and Tableau Online.

## Architecture inside a pod

Each pod is designed to host thousands of customer sites and their users' interactive sessions within a multi-tenant (shared compute) environment. Pods are hosted in AWS, using Amazon Elastic Cloud Computing (EC2) for scalable compute and Amazon Relational Database Service (RDS) for highly available database. Figure 2 shows the pod architecture.



**Figure 2** Architecture inside a pod

Each pod is fronted by an Elastic Load Balancer which distributes traffic to available workers. Where possible, traffic is stateless and sequential requests are routed to available machines. Some traffic (e.g. interactive visualization sessions) is inherently stateful; a visualization session may start on any application worker (chosen by the load balancer) but the interactive visualization is then “sticky” to the machine and process to which it was assigned.

Dual availability zones are supported within each pod to ensure redundant availability.

Durable customer state (e.g. workbooks, data sources, user information, data extracts, bridge configuration) is primarily managed by a storage system that consists of the site and content repository Amazon RDS database as well as cloud hosted network file storage. These systems are backed up and replicated in redundant availability zones. The system is designed for high availability and elastic capacity upgrades; for example, system storage can be quickly expanded with pod growth.

Three kinds of worker machines are managed as pools of Amazon EC2 compute resources. All are designed for easy expansion with pod growth.

- **Application Workers** service requests initiated by end users via browsers, administrative APIs, desktop clients and mobile clients.
- **Background Workers** service scheduled tasks. The pool of background worker machines pulls scheduled work from the Site/Content Repository (Amazon RDS). Background work includes extract refreshes, alerting, and email subscriptions.
- **Bridge Connectors** manage connections initiated by Tableau Bridge clients and route live-query requests to the customer-hosted bridge clients.

These worker machines are kept up to date with the most recently released versions of Tableau's software. Tableau Online hosts many 24/7 mission-critical sites where availability is of utmost importance and can't afford business discontinuity during system upgrades. Most system upgrades and required patches are performed with no downtime, thereby causing no disruptions to the customer's business.

## Security

Tableau has achieved SOC 2 compliance for the Tableau Online service. Tableau services are hosted in data centers that are SOC 2 Type II audited. Copies of these reports are available under NDA. More information and a copy of the Tableau Online SOC 3 report are available [on our website](#).

Automated and manual vulnerability testing is done as a part of the development process and third-party security firms are leveraged to conduct penetration testing of applications before major releases. Quarterly audits are performed for critical elements of the Tableau environment. Learn more about [Tableau Online Security in the Cloud](#).

## Backups

Pods are backed up for disaster recovery purposes. Tableau Online backs up its stateful data for each pod daily. For redundancy, the backups are replicated across multiple Amazon Availability Zones in their AWS Region. Backup retention is 30 days. Tableau Online periodically tests system recovery from backups.

## Resource governance

We are always looking to improve our ability to grow, scale, and manage Tableau Online predictably and efficiently. To provide customers with a stable and reliable experience, Tableau Online has built-in resource governance that limits outlier usage patterns in one customer's site from negatively impacting other customers. Learn more about [Tableau Online site capacity](#).

## **Supporting 24/7 mission-critical sites**

Tableau Online is built for high availability. Tableau takes advantage of both the high availability features available in the product as well as cloud architecture best practices to deliver a reliable experience on Tableau Online. There are many automated monitoring processes as well as engineers on call 24/7 in the event a condition is detected that requires human intervention.

Tableau actively monitors system capacity (e.g. machine processor utilization, background utilization, queue time for background tasks, file input/output, network bandwidth utilization) and has processes in place to add additional worker machines and additional file/network throughput as needed to handle peaks in traffic. Tableau can also isolate demanding workloads and route them to specialist worker machines within a pod. Because all Tableau Online infrastructure is on virtualized cloud infrastructure, we have high resource elasticity that can be used to grow the pods and route traffic to meet demand.

Tableau actively keeps track of the load inside each of our pods and has a healthy engineering factor-of-safety that plans for additional pods before approaching the capacity limits of existing pods. Tableau regularly expands capacity by creating new pods in existing and new regions.

## **Service Level Agreement (SLA)**

Tableau is committed to running Tableau Online with a monthly availability percentage of 99.9% for each regional pod. This percentage is based on the success and error rate of key functions and means our customers will be able to access and explore their site with better than 99.9% availability each month. We measure success by tracking our customer's ability to do things like sign in, access the home page, and successfully navigate and access their projects.

Customers enrolled in Tableau Online Premium Support are eligible for credits in the event of Tableau Online not achieving better than 99.9% monthly availability percentage for a given month. Additional details for this are available in the Service Level Objective and Service Level Agreement (SLA) sections of the Tableau Online Support Policy and Tableau Online Premium Support Policy respectively—[read them on our website](#).

## **Data Management and Resource Blocks**

With the Data Management Add-on, Tableau Online enables customers to run scheduled Tableau Prep flows to combine data from a variety of sources, transform it, clean it, and output high quality published data sources. Because data transformation jobs can be long-running and resource-intensive processes, and often need to run at definite times to serve customers' daily needs, Tableau Online offers the capacity to customers through purchased Resource Blocks. By default, the Data Management Add-On for Online comes with one Resource Block, allowing Prep Flows to be run sequentially throughout the day. Customers needing more concurrency can purchase additional Resource Blocks. Through Resource Blocks, Tableau Online supports both common data transformation needs and can scale up to demanding customer needs that require many concurrent data transformation jobs running 24/7.

## **Proof points: Worldwide customer usage**

Since launching in 2013, Tableau Online now serves over 19,000 customer sites with 450,000+ seats. (customers range from 1-10 seats to over 12,000 seats). From data gathered for the period between January and March 2020, Tableau Online's worldwide pods serve the following customer needs.

- **19,000+ customer sites**
- **14,000,000+ views/month**
- **1,400,000+ email subscriptions/alerts monthly**
- **700,000+ workbooks (over 1,600 customer sites have more than 100 workbooks each)**
- **2,000,000+ data connections**
- **280,000+ extract refreshes daily**



# Case study: Tableau’s alpo-dev site on Tableau Online

Tableau has two major deployments to service its own business and operational analytics needs. One is a company-wide Tableau Sever deployment known as “alpo,” and the other is a departmental Tableau Online site that supports the core business activities of the Development team known as “alpo-dev.”

Alpo-dev runs on a Tableau Online pod and is supported by the same team and engineering operations processes as customer pods. Workbooks on this site are business-critical to Tableau’s development team. The vizzes on automated test results, status of the continuous development pipeline, defects, and Tableau Online usage are required to carry out engineering activities on a day-to-day basis.

Members of Tableau’s development team are encouraged to actively use alpo-dev, and many people publish content daily, ranging from critical product metrics to data reporting the availability of free food left over from meetings in Tableau’s Seattle headquarters. These vizzes are powered by a variety of data sources, including on-premises, Tableau-managed on AWS, and in the public cloud. As such, Tableau’s own use of Tableau Online serves as an excellent case study for rich enterprise cloud customer usage.

Figure 3 below illustrates the data topology.

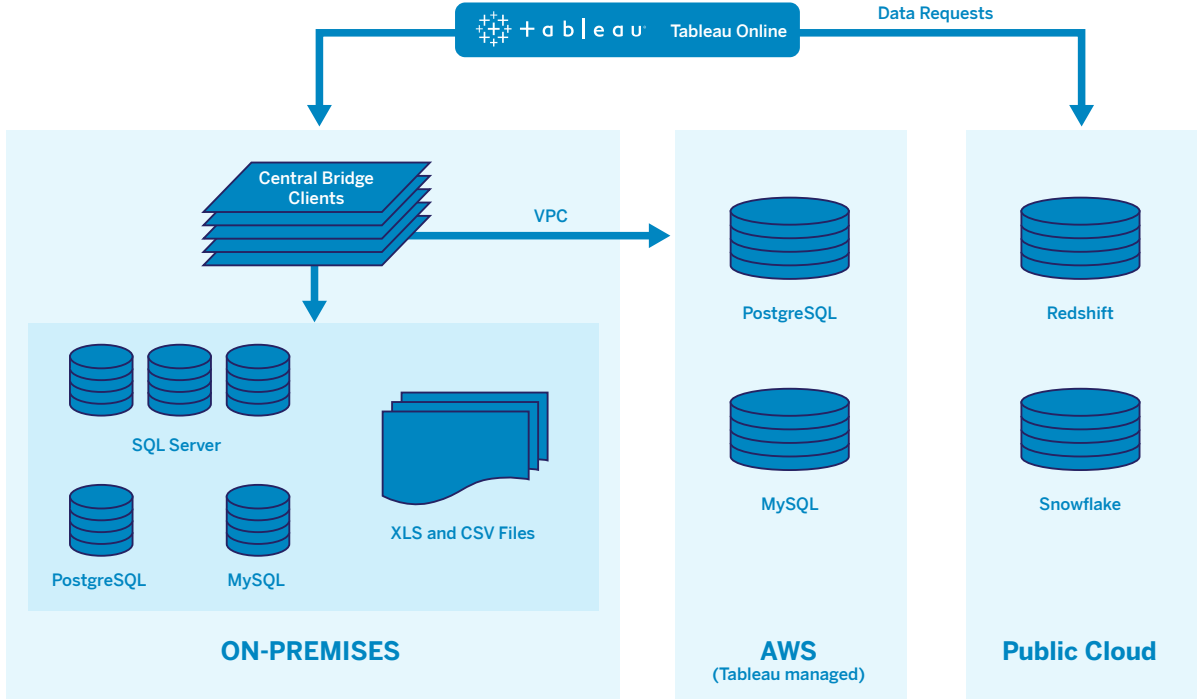


Figure 3 Data connectivity on alpo-dev

## How does alpo-dev resemble typical customer enterprise sites?

- **Enterprise identity** — Identity management and authentication are managed with an enterprise identity provider (IdP).
- **On-premises and cloud data** — There is a hybrid data architecture with a mix of servers including on-premises, Tableau-managed in AWS, and public cloud. There are many different data sources, including SQL Server, PostgreSQL, MySQL, Redshift, Snowflake, Google Big Query, and flat files. A pool of Tableau Bridge services facilitates the live on-premises or VPC database query capabilities, as well as scheduled extract refreshes.
- **Worldwide usage** — Users are located around the world, and view Tableau Online through the web, desktop, and mobile clients inside the corporate network, through the VPN, and over the public internet.

## How is alpo-dev different from typical customer Online usage?

- **Advanced deployment** — Tableau runs pre-release software on alpo-dev's pod. We do this to ensure changes to Tableau Online's software and infrastructure work at scale before we deploy to customers.
- **Heavy usage patterns** — Our users are extremely active! One of Tableau's company values is "We use our products," and alpo-dev reflects this value. As you might expect, Tableau's own internal usage patterns represent above-average usage.
- **All users are Tableau Creators** — Any user can create and publish a workbook or data source. This means the alpo-dev site has lots of workbooks and lots of extracts (both centrally managed and personal).
- **Heavier workload** — Because of our very active user base of creators, the alpo-dev workload is heavier than the workload of a typical enterprise site with a similar number of provisioned users.

## Site configuration

Alpo-dev runs in the US West Coast region. The US West Coast region was chosen because the largest number of Tableau's engineers and many data sources reside in this region, although the site does have regular users from the US east coast and Europe.

## Authentication, user, and group provisioning

Like many companies moving to the cloud, Tableau has adopted a cloud-based Identity Provider (IdP)—in our case Okta—that provides single sign-on (SSO) capabilities to the many applications deployed at Tableau as an enterprise standard. Users can also be provisioned using the IdP, and the IdP also manages security requirements such as two-factor authentication (2FA).

The alpo-dev site is configured to use Okta as its SAML provider for authentication, so users can use their regular corporate credentials to authenticate to the site. This IdP is configured to require 2FA to connect to Tableau Online outside of Tableau's network and VPN. SCIM is also enabled on alpo-dev, so Okta can provision and de-provision users through Tableau's own internally managed IT groups.

## **Configuring and provisioning Tableau Bridge**

Tableau relies on many on-premises data sources and data sources within VPCs, including SQL Server, PostgreSQL, and AWS Redshift data sources. Tableau Bridge is an important part of the alpo-dev deployment. Published data sources connect live to on-premises data and extracts taken by Tableau Bridge.

For increased reliability and load balancing, alpo-dev administrators maintain a central pool of Tableau Bridge clients for live connections and extract jobs. Here are some details about the configuration:

- After monitoring extract times and failure rates using the provided administrative views, administrators chose to run eight bridge clients.
- All bridge clients run in a single pool, and there is no differentiation between clients for live and extract jobs.
- For IT administration and reliability, Tableau Bridge runs in service mode, so the client runs in the background and is automatically restarted at reboot. There is a shared AD account that owns the clients.
- No special firewall configuration is required. Bridge clients sign in to Tableau Online and authenticate with the supplied credentials in the same way as any other web application.

End users can also use their own Tableau Bridge clients in the alpo-dev environment to manage extracts. Users contact the site administrator if they would like for their extract to be managed centrally. Additional details on managing and running Tableau Bridge are available in our [Help documentation](#).

### Monitoring and troubleshooting the alpo-dev site

Alpo-dev has a designated site administrator. Administering the alpo-dev site is a part-time set of duties because the overall system monitoring is handled by Tableau Online.

There is monitoring software installed on the machines hosting the bridge clients, which monitors various metrics including uptime, connectivity, hardware utilization, etc.; no other software is required.

Tableau Bridge analytics can also be seen in the site's administrative reports page. The Tableau Bridge settings page includes status information on each bridge client, which can help troubleshoot issues. The administrative views included in the Status page provide insights needed to investigate other site issues reported by users. No specialized software is used to troubleshoot routine analytics issues.

### Alpo-dev site statistics

Alpo-dev has a very engaged user base. In a seven-day range, nearly 900 of over 1400 provisioned users were active on alpo-dev. On a single day basis, we may see 500-600 active users.

Like most sites, alpo-dev has peaks and valleys of traffic. Most workdays see over 3000 views a day. A recent meeting resulted in a spike of traffic of over 7000 views in a single day on alpo-dev without any degradation in site performance.



Figure 4 View count on alpo-dev

Like many companies, there are a few hours of the day where there is a burst in traffic. During these peak hours, we typically serve over 500 views in an hour, though on a recent burst date we saw 1,264 views served in the peak hour.

As we saw in Figure 3, Tableau uses a wide variety of data sources. About 25% of data source accesses were through Tableau Bridge. Tableau sees anywhere from 1700 to 2500 bridge extract refreshes per day. This represents hundreds of GB in data throughput through bridge clients.

## Appendix – Q&A

Here we've compiled some common questions from enterprise customers and our answers.

**Q: This whitepaper explains that Tableau Online now uses elastic file storage. How are my extracts partitioned and indexed? Is there physical separation?**

**A:** The extracts are stored in elastic storage, and each extract is itself indexed uniquely in that storage (somewhat similar to how Amazon S3 offers a large bucket where discrete pieces of information are stored with unique keys). These indexes themselves are logically partitioned per customer. In a cloud hosted service, tenants are logically separated (it is very rare to physically separate the data for customers). Given this, the operative questions are “How good is the logical partition of the data?” and “At what layers is this partitioning enforced?”

The top-level partition is in the database where customer data is logically partitioned by customer tenant ID (essentially a very big number unique to each customer). This partition ID is used to “stripe” all the other resources we manage per customer and enforce in our application. It's important to understand that in **no case whatsoever** can customers directly access our databases or underlying storage systems; all customer access is only through application logic, which enforces customer-tenant partitioning at multiple levels.

All this said, if you require “physical separation,” then cloud hosting is probably not right for you—in these cases Tableau Server is the right choice (e.g. to support HIPAA, PII data, etc.). In almost all cases, when people talk about cloud systems, the partitions are by definition logical partitions.

**Q: My organization wants to be able to use our own external tools to analyze site administrative and auditing information. Can I have access to access log data from Tableau Online?**

**A:** We love the scenario, but access to log file data is not the solution. Tableau Online will absolutely not share logs, as that would be a bad security practice—this is not the intended purpose of logs, which is for system diagnostics—but the auditing need makes sense. Tableau Online supports Online Admin Insights, the ability to build custom workbooks that audit the sites' activity. If there is a need to export this data (e.g. as a CSV for further analysis), you can use tools such as TabCmd.

**Q: Can I set my own backup policy on my Tableau Online site?**

**A:** If you wish to create or download a local copy of your content, you can do so through the Tableau Online APIs. .

**Q: Does Tableau work with third parties to validate and test Tableau Online security?**

**A:** Yes. In addition to an in-house security team, Tableau works with multiple third-party cyber security experts to perform penetration testing and other security related auditing.

**Experience the reliability and scalability  
of Tableau Online — get started today!**

**START YOUR FREE TRIAL**

## About Tableau

Tableau is a complete, integrated, and enterprise-ready visual analytics platform that helps people and organizations become more data driven. Whether on-premises or in the cloud, on Windows or Linux, Tableau leverages your existing technology investments and scales with you as your data environment shifts and grows. Unleash the power of your most valuable assets: your data and your people.

## Additional Resources

[Tableau Online Security in the Cloud](#)

[Tableau Online: Keeping Your Data Fresh](#)

[Help documentation for Tableau Bridge](#)

