



Accessing Your Private Network Data with Tableau Cloud.

Best Practices for Data Connect & Tableau Bridge

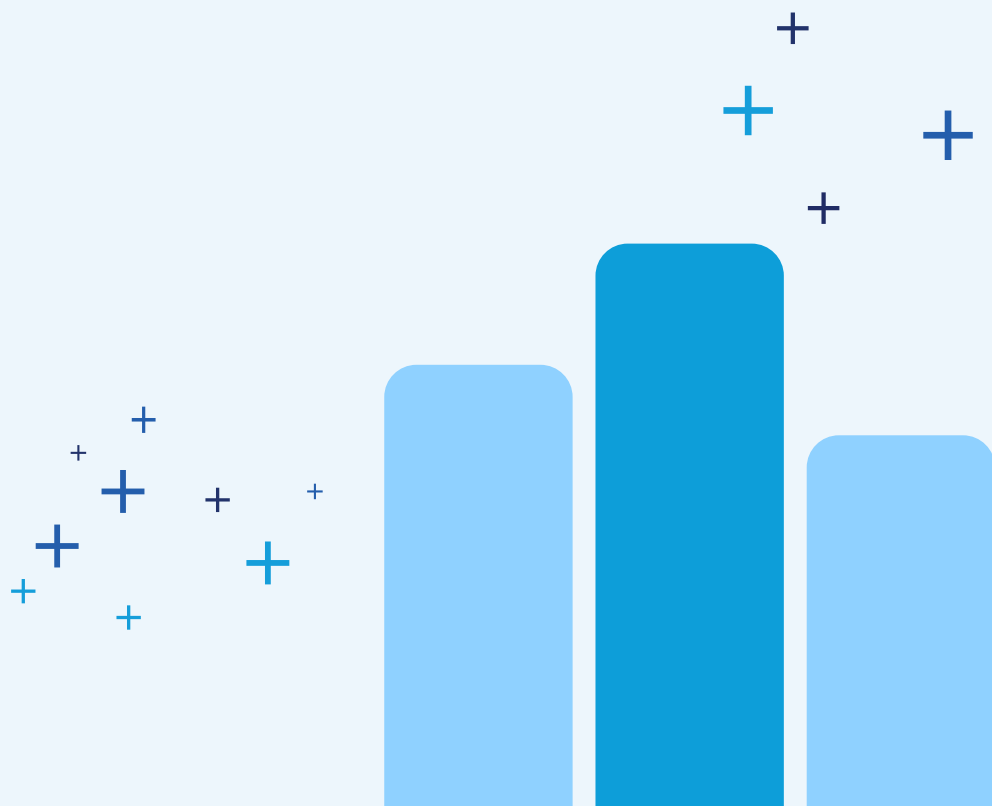
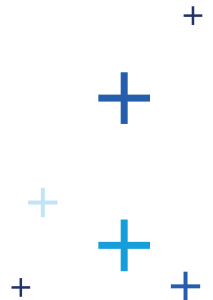


Table of Contents



03 Introduction to Data Connect and Tableau Bridge

- Architecture
- How it works

14 Personas and responsibilities

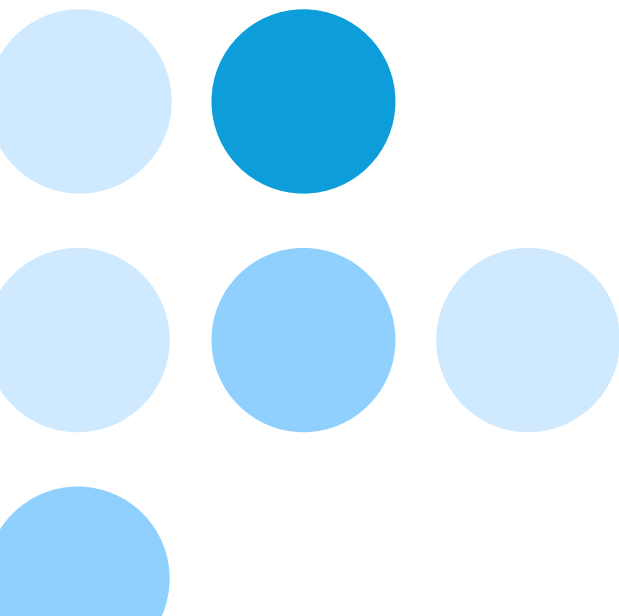
- Administrators
- Data source owners (Creators)
- Workbook owners and analytics consumers (Explorers and Viewers)

23 Deployment scenarios and best practices

- Installation
- High availability
- Scaling considerations

32 FAQ

33 Additional Resources



Chapter One

Introduction to Data Connect and Tableau Bridge

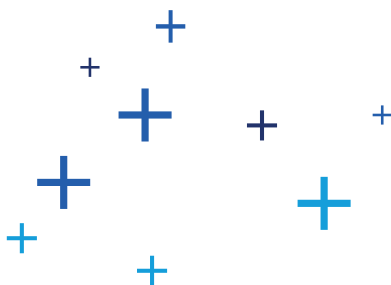
Introduction to Data Connect and Tableau Bridge

[Tableau Cloud](#) is a powerful platform for analyzing data, no matter where the data resides. To realize the value of Tableau Cloud, customers need to unlock all of their data—including data that resides on-premises or in virtual private clouds.

For many years, customers have used Tableau Bridge to access their private data. To simplify deployment and administration of Tableau Bridge, Tableau developed a managed service of Tableau Bridge called Data Connect. This enables seamless scalability and significantly reduces administration of private network data access.

Data Connect is a Tableau-managed Kubernetes cluster running agents (Tableau Bridge clients) on customer-provided infrastructure. Data Connect orchestrates the deployment, upgrading, scaling, and monitoring of the agents across an organization's private network. This allows customers to take full advantage of their private network data without worrying about managing Tableau Bridge themselves.

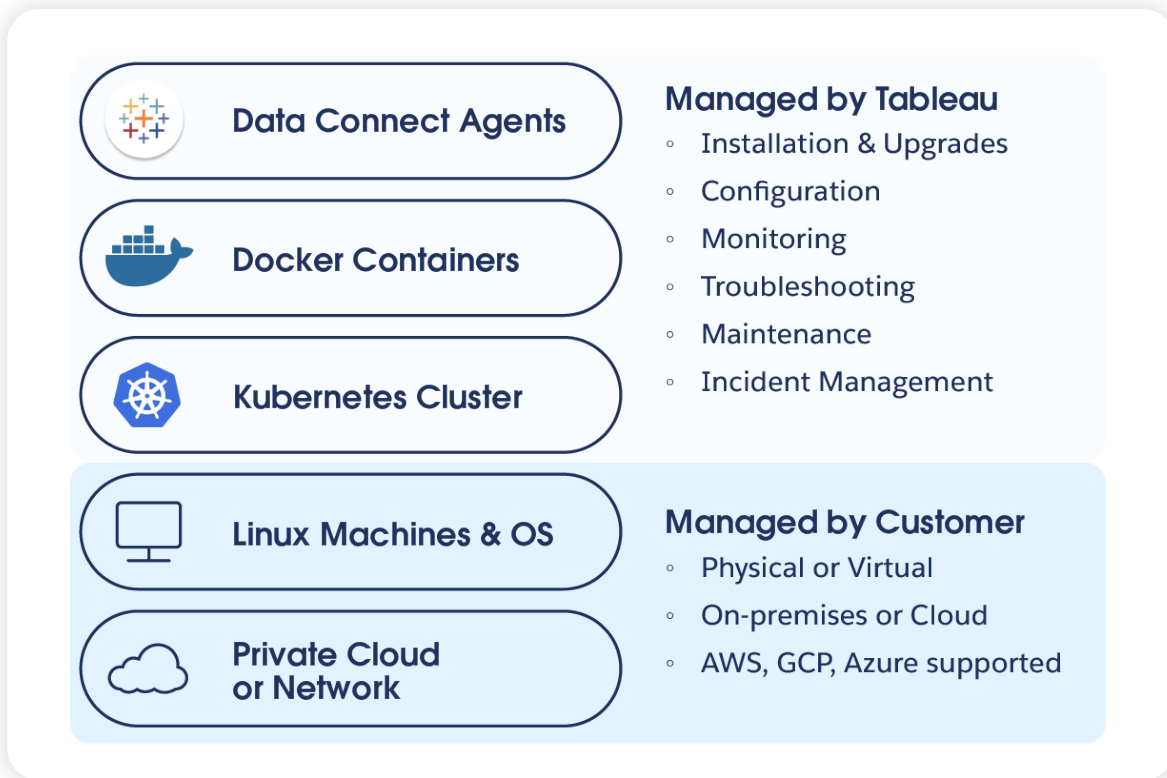
For those that prefer to continue managing it themselves, the Tableau Bridge client is a software application that can be installed on a physical machine or a virtual machine within an organization's on-premises network or virtual private cloud.



Data Connect

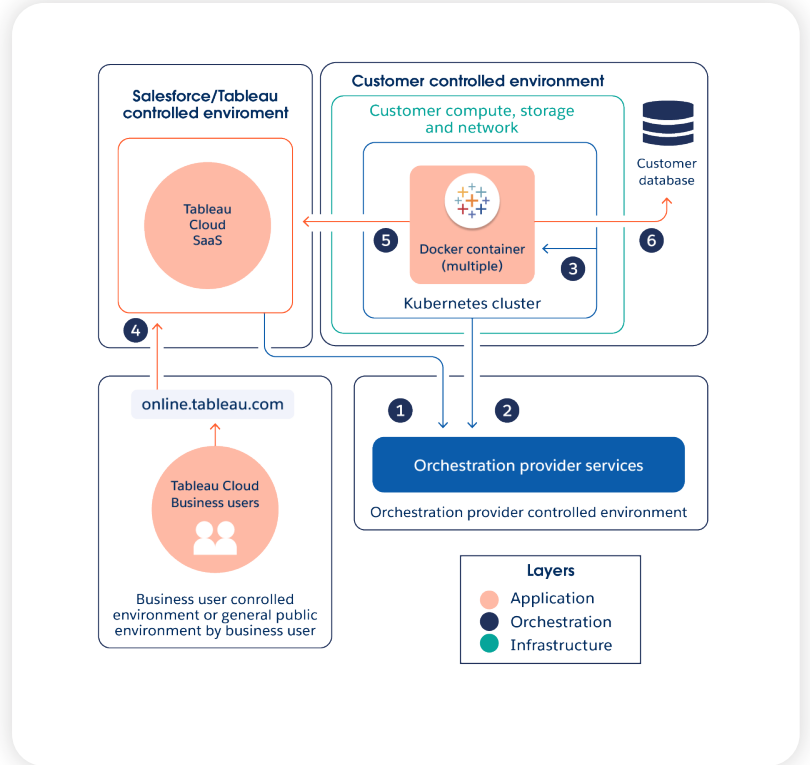
Data Connect operates as a shared responsibility model. With this model, customers supply the physical or virtual compute resources, and Tableau hosts and manages the Data Connect Kubernetes cluster on those resources. Tableau reduces the overhead of administration by remotely managing, monitoring, and maintaining the Kubernetes cluster. In this model, Tableau is responsible for operating the Data Connect service securely and customers are responsible for managing the infrastructure and networking layers.

The Data Connect architecture consists of three main components and responsibility boundaries. Below is an overview of those components and their underlying architecture. While there is some overlap, Tableau is primarily responsible for the application and orchestration layers and customers are responsible for the infrastructure layer (compute, OS, networking, and storage).



Data Connect

1. Tableau Cloud communicates with the Kubernetes orchestration service to deploy, monitor and manage the Kubernetes orchestration.
2. When a customer first initializes Data Connect, a secure connection is established with the orchestration provider service over port 443.
3. Once the service is configured, a Kubernetes cluster deploys a container(s) with Data Connect agent(s). These agents will be responsible for executing Tableau workloads.
 - a. A single Data Connect agent is a Tableau Bridge for Linux for Containers client.
4. Tableau Cloud users log into Tableau Cloud to interact with the Data Connect service.
5. On setup, Data Connect agents initialize a connection with Tableau Cloud using HTTPS. After successful connection, Data Connect agents initiate secure, bidirectional communication with your Tableau Cloud environment using a WebSocket (wss://) connection.
6. Queries initiated from Tableau Cloud will be executed against a customer's database to support end user analysis.



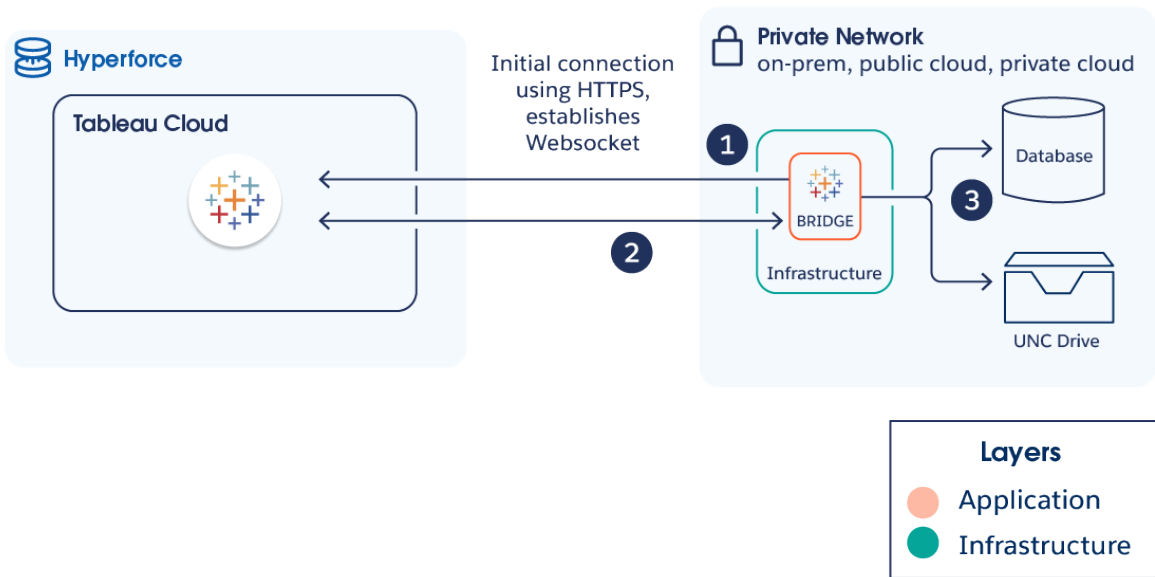
With the ability to perform remedial actions to enable continuous availability, Tableau eliminates the need to monitor traffic and connection status. In addition, to reduce latency and lower network congestion, Data Connect allows customers to determine the data center, edge locations and environments that best meet your requirements for performance.





Tableau Bridge

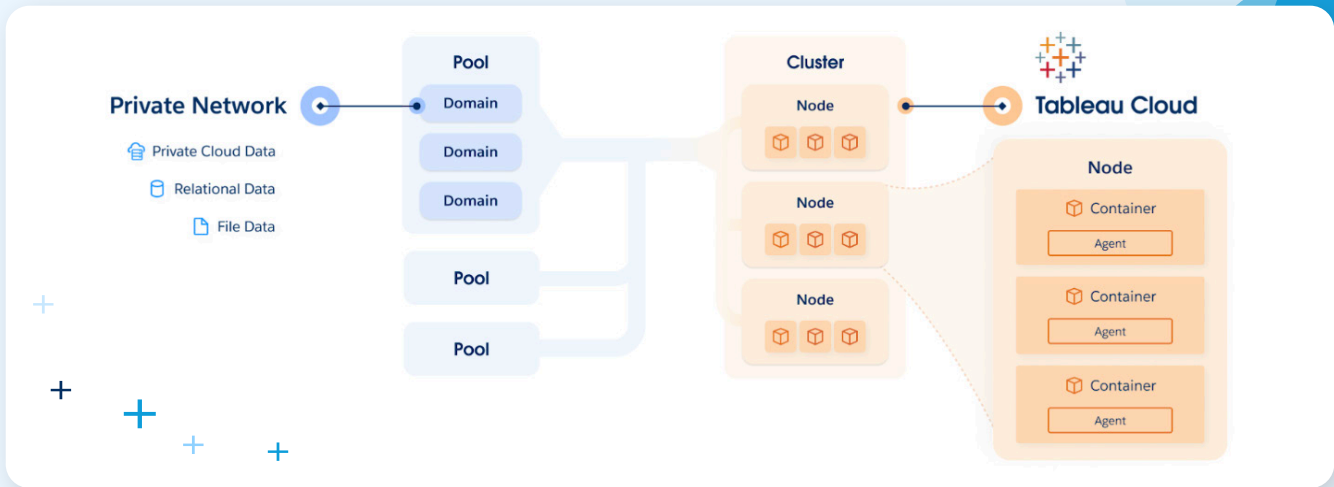
Tableau Bridge is a client software application that runs on a machine in a customer’s network. Similar to Data Connect, it includes both an application layer and an infrastructure layer. It does not include an orchestration layer by default. Tableau provides the application to customers to manage and deploy on their own infrastructure. Unlike Data Connect, Tableau Bridge is entirely managed by the customer.



1. On setup, Tableau Bridge initializes a connection with Tableau Cloud using HTTPS.
2. Once successfully configured, Tableau Bridge initiates secure, bidirectional communication with your Tableau Cloud environment using a WebSocket (`wss://`) connection.
3. Queries initiated from Tableau Cloud will be executed against a customer’s database to support end user analysis.



How it works:



The primary component of the Data Connect system is a cluster. The cluster is a Kubernetes cluster consisting of one or more nodes. Each node is running at least one container with a Data Connect agent. Multiple nodes are required within each cluster to maintain service availability. Data Connect agents are responsible for fulfilling queries against your private network data.

Pools are assigned to clusters to distribute load across the available agents within the cluster. A pool is a logical grouping of domains that specify which clusters should complete specific queries. In the context of deployment planning, a pool is a lookup table containing a collection of endpoints (domains or IP addresses) for the purposes of load balancing. These domains map to private cloud data, relational data, file data, etc.

Clusters

A cluster refers to a Kubernetes cluster that is managed by Tableau. The cluster is the central coordinator for Data Connect, allowing central management and orchestration of all nodes and agents. Action taken on the cluster will impact all nodes and agents within the cluster.

Customers are responsible for providing a base image to the cluster that specifies the drivers and pool utilized by the cluster. The cluster will deploy that base image to all agents on all nodes in the cluster. This allows the service to scale in a harmonized way as the number of agents and nodes change.



Nodes

A node is a single machine running Data Connect within a cluster. Nodes are physical or virtual machines that are provisioned by the customer. They can be running either on-prem or in a private cloud. A single node will run one or more containers containing Data Connect agents at any given time. To maximize capacity of the service, Tableau manages the number of agents deployed on a node and their configurations.

Tableau recommends having at least three nodes available per cluster to maintain availability of the service. There is no licensing limit to how much CPU or memory you have on a node. In most cases, capacity of the underlying machine will be the limiting factor as it relates to performance. Tableau provides machine utilization data and alerts to keep admins up to date on machine capacity within the service.

Clusters and nodes are deployed within a single private network and cannot span isolated private networks; however, customers may run multiple clusters in order to access data in different private networks. There must be at least one cluster deployed for each isolated private network that data will be accessed from.

Data Connect agent

A Data Connect agent is a single deployment of a Tableau Bridge for Linux for Containers client. These agents are deployed onto nodes and managed by the cluster. The cluster will deploy base images provided by the customer to all agents in the cluster. Individual agents are responsible for fulfilling customer queries (live and extract).

Pools

A pool is a logical grouping of domains or IPs that Data Connect clusters will service. Pools allow customers to assign specific clusters to handle particular domains. This gives customers the ability to manage data access at the network level. It also defines how load balancing of data refresh tasks will work.

A common customer scenario involves a dedicated private network for a satellite office. The customer first designates the network's domain to a specific pool. When an extract refresh for a data source is initiated in this network, Data Connect maps the task to that pool using the domain's value. The designated cluster services the pool and then assigns the task to an available agent within the cluster. This ensures resources are efficiently managed and optimal performance is maintained for network tasks.

Self-managed Tableau Bridge

Tableau Bridge deployments can be also managed directly by the customer. Tableau Bridge is client software that runs on a machine in your network. The client works in conjunction with Tableau Cloud to keep data sources that connect to private network data, which Tableau Cloud can't reach directly, up to date.

Individual clients are provisioned on machines that can be deployed either on-premises or in a virtual private cloud. Each client will rely on the hardware resources of the underlying machine. Tableau Bridge client health, maintenance, and scalability are all managed by the customer when you opt to self-deploy the offering. Tableau Bridge was designed to be implemented by a Tableau Cloud Site Administrator on behalf of the users on their Tableau Cloud Site. This provides a centralized point of administration to ensure the correct data is available to consumers and only appropriate data is sent to Tableau Cloud.

Just like Data Connect, self-managed Tableau Bridge leverages pools to distribute workloads across multiple Bridge clients. Domains are mapped to individual pools and pools are mapped to one or more clients. Each Tableau Bridge client must be configured individually.

Private network workload types

Support	Published Data Sources (PDS)	Embedded Data Sources (EDS)	Virtual Connections (VConns)	Prep Conductor Flows
Supported by Data Connect and Tableau Bridge	✓	✓	✓	Not Supported (as of 24.3)

Private network workload types

Below is a brief description of each of the workload types that are relevant for private data connectivity.

Published Data Sources

When you are ready to make a data source available to other users, you can [publish it directly to Tableau Cloud](#). This allows other users to leverage this data source as a single source of truth across places of analysis. Published Data Sources can be either live or extract connections.

Embedded Data Sources

When a workbook is published to Tableau Cloud and packages the data source connection and any metadata provided directly within the workbook. These data sources cannot be shared by other assets. Embedded Data Sources can be either live or extract connections.

Virtual Connections

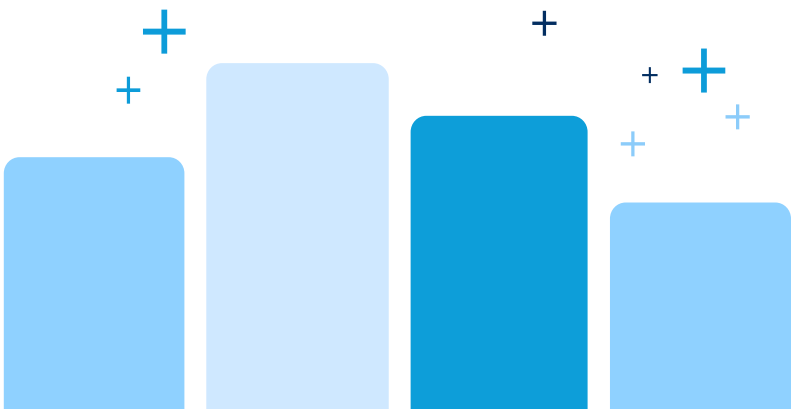
A sharable resource that provides a central access point to data. A virtual connection can access multiple tables across several databases. Virtual Connections let you manage extracting the data and the security in one place, at the connection level. Tables within Virtual Connections can be either live or extract connections.

Prep Conductor Flows

Customers can schedule Prep Flows to update on a scheduled basis and monitor the history of those jobs.

Query execution

In order to support different customer scenarios, Tableau has designed Data Connect and Tableau Bridge to handle Published Data Source extracts differently than non-Published Data Sources extracts. In this section, we will cover the differences between the two options and provide example scenarios of when it makes sense to leverage the strengths of a each option.



Comparison

Below is a summary of the differences between the two connection options.

	Features	Best for	Limitations	Monitoring
Published Data Source (PDS) extract	Creates extract locally on Bridge client	Better performance when creating larger extracts	Single network support	Jobs are surfaced as Bridge jobs in monitoring solutions
All other scenarios (PDS Live, EDS, VConn)	Data sent from Bridge clients to Tableau Cloud to service the query Extract only: Extracts created on Tableau Cloud	Cross-network queries Offloading extract creation to Tableau Cloud	Subject to Tableau Cloud's Site capacity limitations and query timeouts Larger queries are less performant	Jobs are surfaced as Cloud Backgrounder jobs in monitoring solutions

In order to support different customer scenarios, Tableau has designed Data Connect and Tableau Bridge to handle Published Data Source extracts differently than non-Published Data Sources extracts. In this section, we will cover the differences between the two options and provide example scenarios of when it makes sense to leverage the strengths of each option.

Performance

From internal testing, Published Data Source extracts are typically faster when compared to Embedded Data Source extracts and Virtual Connection extracts using Bridge. While results may vary, our internal testing indicates that the performance difference expands for larger workloads. This is due to the fact that the extract is created locally and then sent to Tableau Cloud. For more information about how to manage performance differences between the two connection types, head to [Optimize Bridge Refresh Performance](#).



Cross-network Queries

For non-Published Data Source extracts, a query serviced by private network data solutions can span network boundaries. Since these queries aggregate results centrally in Tableau Cloud, they are able to receive responses from multiple sources to build the overall result (live or extract). This allows Data Connect and Bridge customers using non-Published Data Source extracts to join:

- Public data with private network data
- Private network data with data from another private network
- Data from multiple public and private networks

For Published Data Source extracts, the query cannot span network boundaries. The Data Connect agent or Bridge client responsible for the query must be able to communicate with every database relevant to the Published Data Source extract.

Monitoring

As each option services queries in different locations, their monitoring solutions are different. A full breakdown of how to monitor these workloads is available in the [Monitoring section](#).

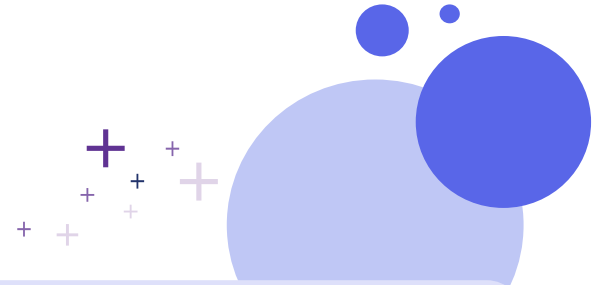


Chapter Two

Personas and Responsibilities

Administrators

Data Connect administration



Data Connect Setup

Data Connect was designed to reduce the burden on administrators and allow the analysis of private data to be seamless and scalable. Admins will be responsible for configuring their network policies, provisioning the infrastructure, and providing a base image to deploy. Once the service is running, administrators' responsibilities entail monitoring the health of the underlying machines, adding nodes as usage increases and rotating base images if database utilization changes, saving time for more impactful activities for your organization. You can find more detail on monitoring system health and scalability [here](#). Individual jobs serviced by Data Connect are also available for monitoring, which is detailed [here](#).

Base Image Creation and Rotation

On setup, administrators will provide a Base Image that specifies how Data Connect agents are configured for a cluster. The Base Image includes the drivers you will utilize and the Pool associated with that Base Image. Tableau will then deploy that Base Image to all agents within the cluster assigned to that pool. This allows the service to scale in a harmonized way across your deployment, even if you add or remove nodes in the future. You are able to make changes to your base image at any time. These changes will be rolled out to your deployment by Tableau on your behalf.



Tableau Bridge administration

Tableau Bridge administrators are responsible for planning, implementing, maintaining, and monitoring the Bridge deployment.

Planning the Bridge deployment centers around determining the number of Bridge clients required to support a customer's workload. Head to [Bridge scaling considerations](#) to learn how to plan your Bridge deployment.

Once a customer has planned their deployment, they must provision hardware (physical or virtual) to host Bridge. Tableau's deployment guide offers recommendations on hardware specifications for both [Windows and Linux deployments](#). Once the hardware is provisioned, admins must install the Bridge software, along with the required [database drivers](#), on that hardware.

Running [Tableau Bridge in a container](#) is only supported on Bridge for Linux. With Bridge installed, admins [configure pools](#) to route data freshness tasks to certain Bridge clients based on the underlying network locations of the data sources.

Now that the Bridge is running, the admin is responsible for ongoing maintenance and monitoring. Maintenance involves upgrading the software when new Bridge releases or patches are released and monitoring performance as usage scales. Monitoring involves tracking both the performance of data refresh tasks, as well as machine-level monitoring of the hardware running Bridge. These items are discussed in more detail in [Bridge scaling considerations](#).

Monitoring your deployment

Monitoring solutions

Several monitoring solutions are available in Tableau Cloud to help you manage different monitoring scenarios for Data Connect and Tableau Bridge. All of these solutions are available to administrators only by default. Admin Insights and Activity Log can be made available for other individuals if desired.

Real time

- [Jobs Page](#): Monitor real-time extract job status for jobs performed in the last 24 hours.
- [Admin Views](#): Monitor real-time extract job status and performance. Includes recent job history for longer term analysis.

Historical analysis

- [Admin Insights](#): Monitor job status and performance using Tableau Cloud data sources that are published in your environment. This data is updated daily and shows historical data for longer term analysis. It can be shared with non-admin users if desired.
- [Activity Log](#): Monitor job status and performance using log data that can be stored as long as necessary for your analysis. This data can be shared with non-admin users if desired. Available for Tableau Cloud Enterprise and Tableau+ customers only.

Real-time monitoring for individual extract jobs

Different workload types result in jobs that are completed either entirely by a Data Connect agent/Bridge client, or that are completed by a [Tableau Cloud Backgrounder](#). Published Data Source extracts are created by a Data Connect agent or Bridge client. Embedded Data Source extracts and Virtual Connection extracts are created by the Tableau Cloud Backgrounder.

Below are workflows available to monitor individual jobs real time for each workload type.

Published Data Source extracts:

From a monitoring perspective, there are two components to a private network published data source extract handled by Data Connect or Tableau Bridge. The first (1) is for Tableau Cloud to hand off the job completely to a valid client. The second (2) is for the client to complete the job, and publish the packaged data source, complete with data, to Tableau Cloud. Below are the ways to monitor each of these components in the Jobs page and in Admin Views:

Jobs page

This page will list the published data source extract Task Type as “Bridge Refresh.”

1. The job will show as “In Progress” until the job is sent to the client.
2. The job will show as “Sent to Bridge” once the job is sent to the client.

Admin Views

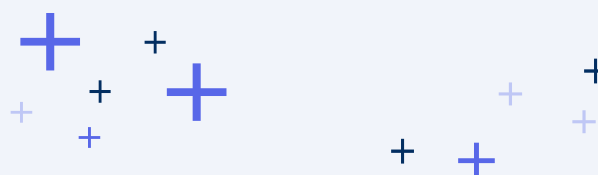
Each component of a Published Data Source extract job is shown in a different view.

1. The Background Task for Non Extracts Admin View, when filtered by the “Task” filter of “Refresh Extracts via Bridge,” will show everything related to sending the job to the agent or client. Since it is not the direct creation of an extract, it will appear on this Admin View.
2. The Bridge Extracts Admin View will list all of the published data source events related to completing the job after it is sent to the client.

Embedded Data Source and Virtual Connection extracts:

The flow of an embedded data source or virtual connection for extracts differs from that of a published data source. In the case of an Embedded Data Source/Virtual Connection extract, the Cloud Backgrounder will stay engaged throughout the entire extract creation and will create the extract using the Cloud Backgrounder.

Below are the ways to monitor the same two components in the Jobs page and in Admin Views:



Jobs page

This page will list the Task Type as “Extract Refresh/Creation.”

1. The job will show as “Pending” while the job is held in the Tableau Cloud Backgrounder queue until the job begins.
2. Once the job begins, it will show as “In Progress.” It will not show as “Sent to Bridge” because the extract will be created on a Tableau Cloud Backgrounder and there may be more than one agent or client involved in the creation of the extract.
3. Once the job completes, it will show as “Completed” or “Failed.”

Admin Views

The two components of the extract will show in different Admin Views.

1. The Background Task for Non Extracts Admin View, when filtered by the “Task” filter of “Refresh Extracts” will display the tasks for embedded data source extracts and virtual connection extracts. This view will show the component of a job is related specifically to sending the job to the agent or client. Since it is not the direct creation of an extract, it will appear here.
2. When the Cloud Backgrounder has completed the work of creating the extract, it will communicate the result of the job in the Background Tasks for Extracts Admin View.

Historical Analysis

Admin Insights and Activity Log enable lookback analysis on the jobs that require Data Connect or Tableau Bridge. The Job Performance data source is the primary data source for monitoring private network data jobs in Admin Insights. It also includes data for other user-facing tasks such as subscriptions that are serviced by the Cloud Backgrounder. Both Admin Insights and Activity Log can also be used as a source to curate your own admin data source by combining it with organizational data or data from any of the other admin data sources.

Example scenarios for both include:

- Performance trends over time
- Variation in extract performance duration over time
- Capacity planning



Data Source Owners (Creators)

Both Data Connect and Tableau Bridge are largely abstracted from analysts and data source owners. In many cases, they will not even realize that these products are servicing their queries. However, there are three areas where they will interact with Data Connect or Tableau Bridge: initial database onboarding, data source configuration, and troubleshooting data source errors.

Initial Database Onboarding

When a new private network database is first connected to Tableau Cloud, a Tableau Cloud Site Admin will need to add the database domain to a preferred pool. This is a one-time effort that requires a data source owner to communicate the database domain with their Site Admin. Please note that this is a database-level configuration, not a data-source level configuration, so it should only take place once per database.



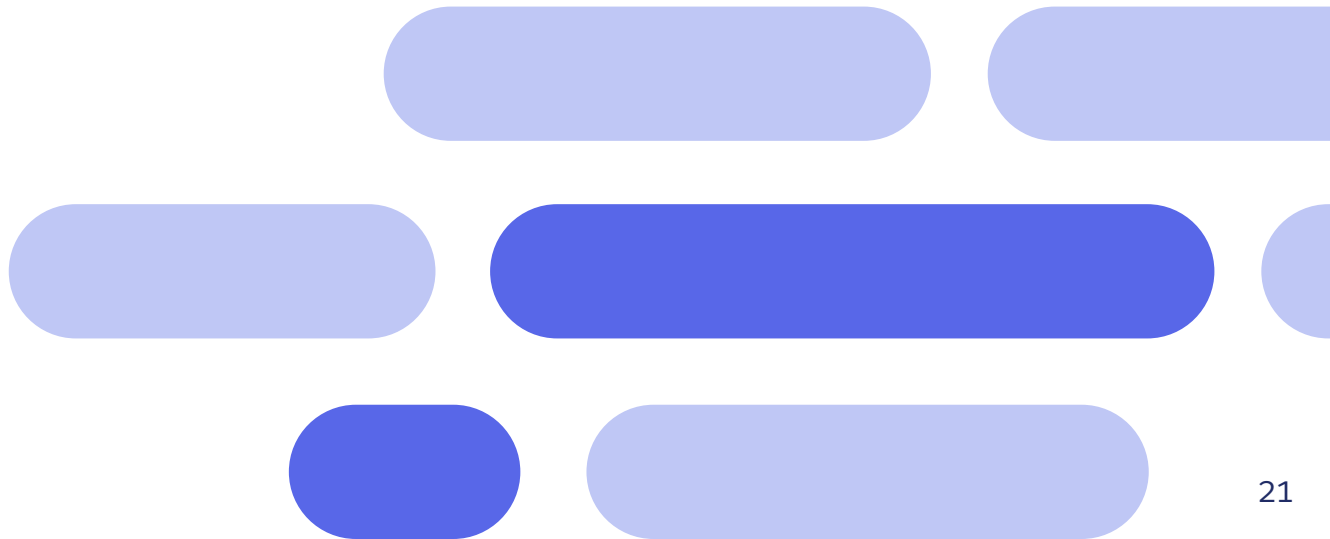
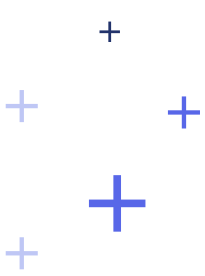
Data Source Configuration

(Published Data Sources only)

Published Data Sources that are built off of Private Cloud data require analysts and data source owners to make an additional [configuration](#) for that data source to leverage Data Connect or Tableau Bridge. Selecting “Private Network” will send that query to Data Connect or Tableau Bridge.

If they select “Tableau Cloud”, the query will be attempted against the database directly from Tableau Cloud without using Data Connect or Tableau Bridge.

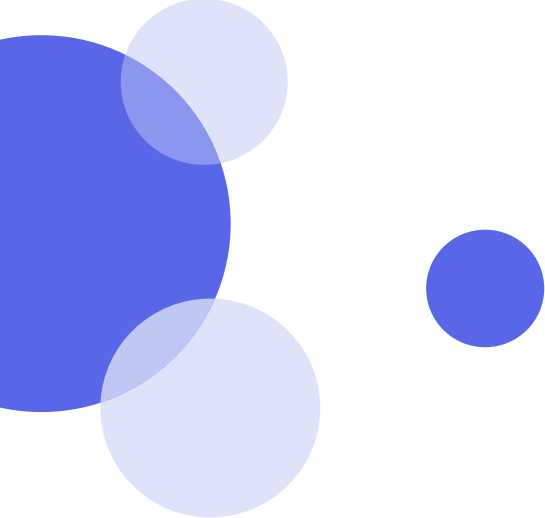
All other workload types do not require this toggle to be selected. Instead, they will rely entirely on the domain of the database to determine if that query should be routed through Data Connect or Tableau Bridge.






Troubleshooting data source extract errors

For any data source extract (Published or Embedded) error that occurs, the data source owner will be notified of the error and provided a summary of the error message. Some errors that occur may involve the health of Data Connect or Tableau Bridge. In that case, the data source owner should notify their Tableau Cloud Site Admin of the issue. Individual analysts are not typically responsible for resolving these issues.



Workbook owners and analytics consumers (Explorers and Viewers)

Both Data Connect and Tableau Bridge are invisible to workbook creators and analytics consumers, meaning these users will likely not be aware that Data Connect or Tableau Bridge are being leveraged. If a user connects to data in a private network, Tableau Cloud will handle the connection and data freshness tasks with no action required from the user. This allows for a consistent and seamless end-user experience.



Chapter Three

Deployment
scenarios and
best practices

Installation

There is no “one-size-fits-all” method to successfully deploy these products. However, there are best practices that can be applied for deployments. We have outlined a few primary items below.

Networking

Everything below applies to both Data Connect and Tableau Bridge:

- All communication from your infrastructure to Tableau Cloud is initiated from behind your firewall. You do not have to manage additional exceptions.
- The machine hosting the product must be able to communicate with all databases that you expect end users to query.
- The “closer” you deploy these products to your database, the lower any potential latency will be.
- Both products require the ability to make an initial outbound connection using HTTPS. After the initial outbound connection, secure, bidirectional communication to your Tableau Cloud environment is established using a WebSocket (wss://).
- Both products support [forward proxy filtering](#).

Hardware

Both Data Connect and Tableau Bridge are primarily responsible for querying live data and creating Tableau extracts. These jobs are particularly memory intensive on the machine. While CPU can be a factor at larger volumes, allocating sufficient memory to your hardware is essential. Data Connect provides machine health metrics within Tableau Cloud.

Minimum and recommended specifications for Data Connect are available [here](#). Minimum and recommended specifications for Tableau Bridge are available [here](#).



High availability

Data Connect

Tableau Data Connect was designed with high availability in mind. Customers must run a minimum of three nodes per private network to maintain availability of the service. Once deployed, Tableau will manage agent health to maintain service availability.

Customers are responsible for maintaining the operating system and the availability of the machines Data Connect is deployed on. This includes any maintenance or patching required on the machines.

- If any of those activities require a restart of the machine, the Data Connect service will automatically restart and should be available shortly after the machine is available. If you change machines during this process, you will need to re-establish connection with the Data Connect service. You can do so by following the [setup guide](#).
- Data source owners will be alerted of any data source failures that occur at this time.

Tableau Bridge

Tableau Bridge can be configured for high availability, but customers must proactively manage key failure points to ensure uninterrupted service. There are multiple points of failure that customers must manage to successfully keep Tableau Bridge available:

1. Leverage pools to distribute workloads.
2. Multiple (three or more recommended) Bridge clients should be configured with all pools utilized to establish system availability.
3. Individual Bridge clients may become unhealthy, in which case Site Admins will need to take action to resolve the issue.
4. Customers need to upgrade Bridge clients in a way that does not impact their workloads.
 1. Add the new Bridge client to an existing pool.
 2. Confirm that jobs handled by that client are being handled correctly.
 3. Remove the old Bridge clients from all pools. This will prevent any new jobs being sent to that Bridge client.
 4. Wait for all current Bridge jobs to complete or timeout. You will need to wait for as long as the currently configured timeout value. The [default timeout value is 24 hours](#).
 5. Disconnect the old Bridge client from the Tableau Cloud site.

Scaling considerations

When planning to scale a deployment, the considerations for Data Connect and Tableau Bridge are very different. Data Connect scalability is based primarily on scaling the infrastructure required to satisfy your Tableau workloads. Tableau Bridge scalability is based on both the infrastructure required and understanding what type of workloads you will be running.

Data Connect

Data Connect was developed to simplify scaling Tableau workloads behind private networks, and thus scales in a different way than Tableau Bridge. During initialization, Tableau provisions several Data Connect agents on every node that enable each node to satisfy significant concurrent workloads. As your workloads grow, Tableau can optimize the number of Data Connect agents along with other configurations to maximize each node's output. Data Connect is a very memory intensive service and will benefit from having larger amounts of memory on each Data Connect node. As a result, a node's hardware specifications, especially memory, will be the limiting factor for performance.

Monitoring scalability

There is no licensing limitation on scaling up the hardware that services Data Connect, but there are licensing implications for scaling out. For this reason, we often see customers choose to scale the service up, before they scale out. Scaling up particularly helps to increase the performance of jobs (e.g. faster extract creation). Scaling out particularly helps to increase the throughput of the service.

Customers are notified when a node's workload approaches hardware limitations. This allows customers time to scale hardware either up or out. Admins can monitor Data Connect workloads from within Tableau Cloud. In general, we recommend having enough hardware such that your memory and CPU do not consistently surpass 80%.

Scaling up hardware while minimizing downtime

Scaling up your hardware can take place as quickly as you are able to provision the hardware. If you choose to scale your hardware up and are using virtual machines, you may not need to restart the machine. If you do not need to restart the machine, you will not experience any downtime to the Data Connect service.

If you do need to restart a machine, you will experience downtime on that node until the hardware is available and Data Connect has re-deployed agents on the machine. This does not require any human operation to complete and generally takes one to two hours, depending on how long the restart takes. To minimize impact we recommend always leaving two (or if possible three) nodes available at any one time. This ensures that the available nodes are able to continue servicing the private network workload. If you only leave one node available within a cluster at a time, then it becomes a single point of failure for your cluster and increases risk of downtime.

Scaling out hardware while minimizing downtime

Adding new hardware in the form of a new Data Connect node does not disrupt the service. You will simply purchase a new Data Connect node (if necessary), and then follow the standard set up process in our documentation. Once that node is successfully configured within its pool, it will begin taking on new workloads. If you need to purchase additional Data Connect nodes, please work with your Tableau account team.

Can I scale Data Connect down?

Customers can remove Data Connect nodes at any time. However, please keep the following in mind:

- Any jobs running on the node when the machine is turned off will fail. For extracts, data source owners will need to restart those jobs manually or wait until the next scheduled refresh for that job to be routed to a new node. For live queries, an end user will see an error message on their dashboard and need to refresh their dashboard to render the dashboard successfully.
- Re-deploying Data Connect nodes typically takes one to two hours to complete in addition to the time it takes to spin up a new machine.
- Data Connect node licenses are annual subscriptions. Turning the node off does not reduce Data Connect licensing costs.

Using pools to manage scale

Both Data Connect and Tableau Bridge [utilize pools](#) to manage scale across the organization. Individual infrastructure is only utilized by the domains associated with a given pool. Customers can load balance workloads per database by using different pools for different databases.

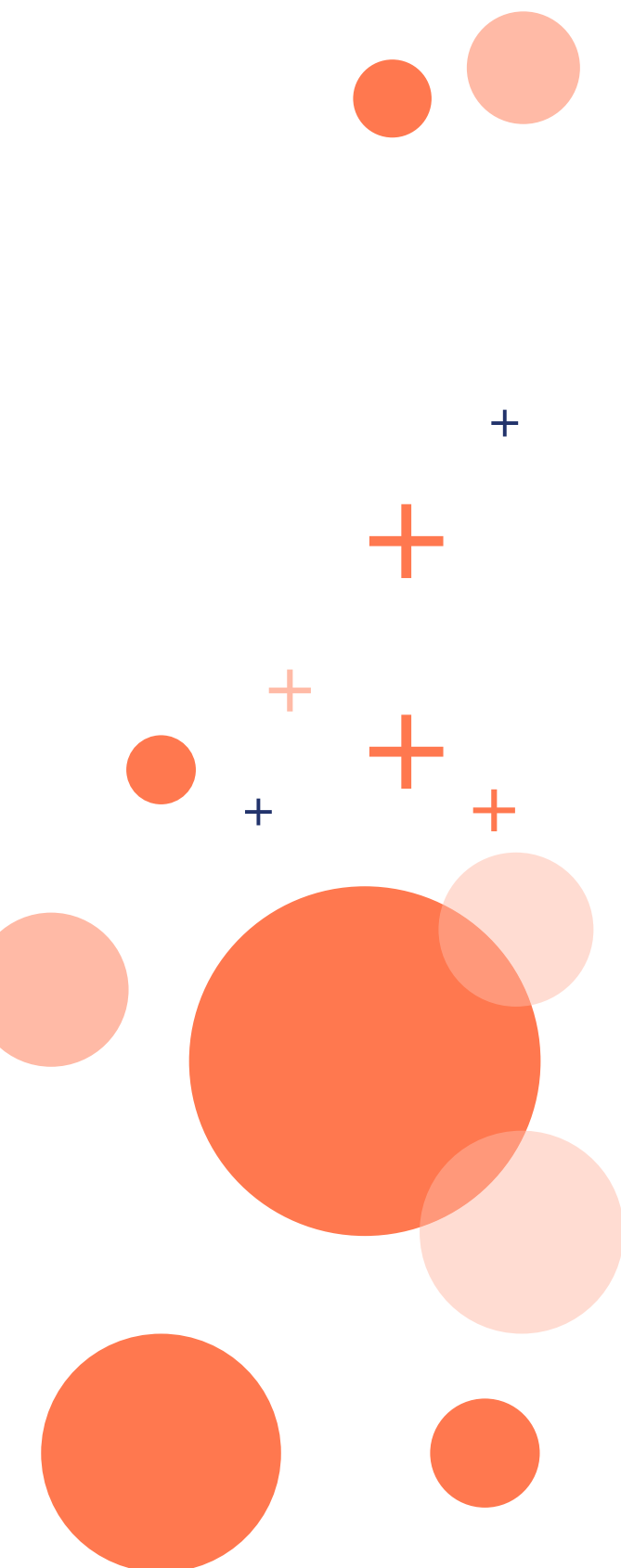
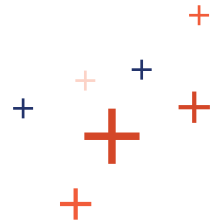


Tableau Bridge

Tableau Bridge scalability is based on both understanding what type of workloads you will be running and the infrastructure required. Customers are responsible for understanding how Tableau Bridge scales and how to monitor the health of self-managed Bridge deployments to maintain acceptable performance levels. Scaling Tableau Bridge up and down is a simple process and has no Tableau licensing implications.

Bridge scalability baselines

- The Bridge for Windows client was designed to have one client installed per machine.
- The Bridge for Linux for Containers client was designed to enable more than one client per machine.
- Bridge clients can only be registered to one site at a time.
- There is no limit on how many clients that can be registered to a site.
- We recommend at least two Bridge clients per pool per site for redundancy.
- Tableau Bridge is designed to scale up and scale out without licensing costs. See [Monitoring Scalability on page 26](#).



Types of workloads and their impact on scale

In practice, different workload types impact Tableau Bridge deployment planning in different ways. We have collected the relevant information below to help customers establish the baselines required to plan their deployment.

Workload type	Primary concurrency limit	Is queued	Customer takeaway
Live query (Published and Embedded Data Sources)	16 queries per Bridge client	Yes	Customers can increase Bridge clients to reduce queue times.
Embedded Data Source and Virtual Connect extracts	10 queries per site*	Yes	Adding Bridge clients does not impact throughput. Spread extract start time out over different times or shift to Incremental Extract Refreshes to reduce queue time.
Published Data Source extracts	Configurable per Bridge client (Default = 10)	No	If the Bridge client is at concurrent capacity, the job will fail. If you see concurrency failures, increase the concurrency configurations or add Bridge clients to the impacted pool.

* For customers with Tableau Cloud Enterprise, [Tableau+](#), or with Advanced Management (now deprecated), the concurrent limit is 25.

Types of workloads and their impact on scale

- Live queries: Tableau Bridge supports 16 concurrent live queries per client. At load time, there will be one query for every data source connection used by a worksheet.
 1. A dashboard may contain multiple worksheets. This means loading a dashboard results in a unique live query for every data source connection on every worksheet on the dashboard.
 2. If all Bridge clients within the pool are saturated, live queries are queued.
- Embedded Data Source extracts and Virtual Connection extracts are subject to the [Tableau Cloud concurrent extract refresh limitations](#) at the site level. For most Tableau Cloud customers, this means that you will never have more than 10 concurrent Embedded extract refreshes, Virtual Connect extract refreshes, or a combination of the two on a site at a given time.
- Published Data Source extracts are not queued. Tableau Bridge supports 10 concurrent extract queries of all types per client by default. This value [can be configured](#) per Bridge client. We suggest customers choose to increase this to 20-30 concurrent extracts to better utilize underlying hardware.
- Concurrent live query and concurrent extract queries do not impact each other's concurrency limitations. In other words, a single client may be working on 16 live queries and 10 extracts at the same time.

Monitoring Tableau Bridge scale

The Job Performance data source helps customers monitor whether a Bridge pool is at capacity for its exact workload. For Embedded and Virtual Connection extracts, the “Overflow Queued At” field in the Job Performance data source of [Admin Insights](#) provides visibility into the queue times for these jobs. Spread extract start time out over different times or shift to [Incremental Extract Refreshes](#) to reduce queue time. To increase the capacity for these jobs, consider upgrading to Tableau Cloud Enterprise or [Tableau+](#). This allows a greater concurrency capacity for Backgrounder extracts.

For Published Data Source extracts, “Tableau Bridge Pool at Capacity” errors are recorded in the Job Performance data source of [Admin Insights](#). This error indicates that the current workload is surpassing its concurrent capacity. To address this issue, either increase the [concurrency settings](#) of those Clients or consider adding more Bridge clients to the pool. In select cases, admins can also opt to work with data source owners to distribute the extract timings they have configured for their data sources.

We recommended that workload-specific concurrency infrastructure capacities stay below 80% to allow for unplanned spikes in demand or natural growth without risking interruption or degradation of services. Customers can use the Job Performance data source in [Admin Insights](#) to monitor the number of jobs per client. Hardware-level monitoring takes place outside of Tableau, using a customer's preferred machine health monitoring solutions to track infrastructure capacity.



Scaling Tableau Bridge up

In general, the memory on a Tableau Bridge machine has the most impact on performance of individual queries. One way that customers can increase performance of individual queries is to increase the memory (and in some cases CPU) of the underlying machines. In most cases, adding memory to a machine will require a restart of the Bridge client. Please follow the steps below to ensure that jobs are not impacted by scaling Tableau Bridge up.

1. Ensure the pool has additional Bridge clients configured so that new jobs are not lost.
2. Remove the Bridge client from all pools. This will prevent any new jobs being sent to that Bridge client.
3. Wait for all current Bridge jobs to complete or timeout. You will need to wait for as long as the currently configured timeout value. The [default timeout value is 24 hours](#).
4. Disconnect the Bridge client from the site.

You can find both minimum and recommended Bridge machine sizes [here](#).

What is the minimum number of Bridge clients required for my workload?

As live queries, Embedded Data Source extracts, and Virtual Connection extracts all allow queries to be queued in some form, they rarely impact minimum client recommendations. Published Data Source extracts do not queue. Published Data Source extracts do not queue. To determine the minimum number of Bridge clients per site, identify the maximum number of concurrent extracts requiring Tableau Bridge, then calculate the minimum number of clients needed. More detail on planning your Bridge deployment per site is available [here](#).

Scaling Tableau Bridge out

In general, increasing the number of Bridge clients increases its maximum concurrent capacity (throughput). When adding a new Bridge client to an existing pool, that Client will immediately begin picking up new jobs that are sent to that Bridge pool. No additional steps are required to scale Tableau Bridge out seamlessly.

Can I scale Tableau Bridge down?

Customers are able to reduce their number of Bridge clients whenever they prefer. There are no Tableau licensing implications. Please note if you remove a Bridge client while it is currently running a job, that job will fail. To not impact production workloads, please follow the steps below to remove a Bridge client:

1. Remove the Bridge client from all pools. This will prevent any new jobs being sent to that Bridge client.
2. Wait for all current Bridge jobs to complete or timeout. You will need to wait for as long as the currently configured timeout value. The [default timeout value is 24 hours](#).
3. Disconnect the Bridge client from the site.

FAQ

Is Tableau Bridge being deprecated with the release of Data Connect?

No. The Tableau Bridge software will continue to be offered as a free download for customers who want to manage Bridge clients on their own.

Can I use Data Connect and Tableau Bridge on the same Site?

Yes. There may be scenarios where a customer wants to use Data Connect for the bulk of their private network data, but they will still use Bridge for one-off private connections. For example, if there is a new private network or short-term need for private connectivity and they prefer to quickly stand up Bridge.

In these cases, the important thing to note is that pools are not shared across the services. That means that Data Connect and Tableau Bridge cannot share domains or IPs that they access on a site.

Where can I learn more about the security of Data Connect?

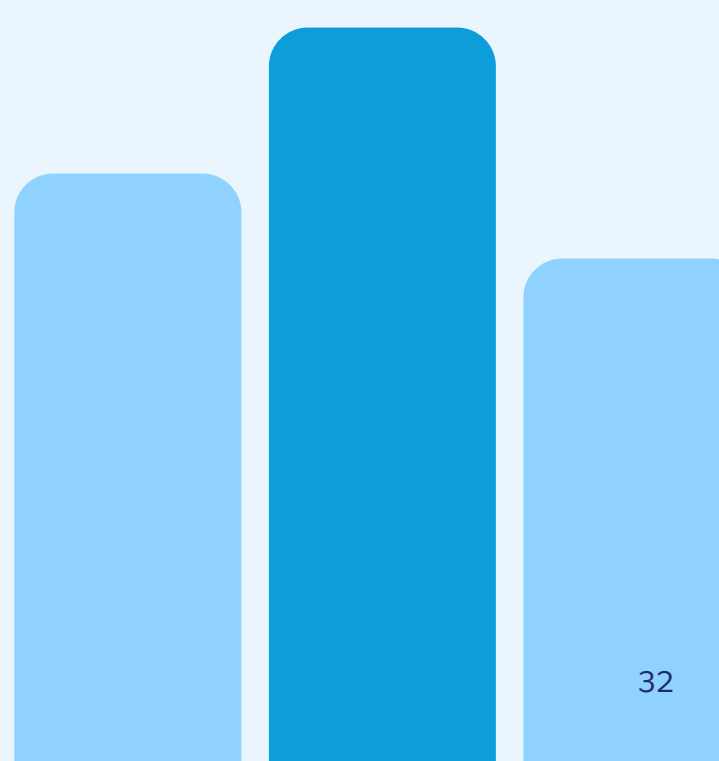
We have detailed all information relevant for Data Connect security [here](#).

Where can I learn more about the security of Tableau Bridge?

We have detailed all information relevant for Tableau Bridge security [here](#).

How do I get started with Data Connect?

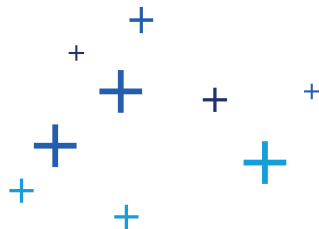
We recommend discussing with your account team to determine if Data Connect is right for you.



About Tableau from Salesforce

Tableau is the world's leading AI-powered analytics platform. Intuitive data experiences, backed by generative and predictive AI capabilities, elevate insights where you work most. Offering a suite of analytics and business intelligence tools, Tableau turns trusted data into actionable insights so you can make better decisions every time. Tableau offers the most choice and flexibility for your architecture as your technology and AI strategy evolve. With security, data governance, and compliance in mind, your organization can maintain agility as new demands on data arise.

Tableau is committed to supporting the unique needs of organizations around the world with the largest partner and success ecosystem, including the passionate Tableau Community that can teach, support, challenge, and celebrate you at every stage of your AI journey. The future is limitless when you start with data and move forward with Tableau. For more information, visit www.tableau.com.



Additional Resources

Want to dive a little deeper?

A number of on-demand sessions from Tableau Conference are available on Salesforce+ that cover Tableau Cloud topics in greater detail:

- [Accessing Your On-Premises and VPC Data with Tableau Cloud](#)
- [Hyper Secure and Scalable: Tableau Cloud on Hyperforce](#)
- [Take Charge in the Cloud with Tableau Cloud Manager](#)
- [Drive Impact: Embrace Trustworthy AI with Tableau Cloud](#)
- [Unlock Seamless Cloud Migration: Mastering the Migration SDK](#)

Not yet a Tableau Cloud customer?

[Start your free trial of Tableau Cloud today.](#)

