+ + + + + + t a b l e a u ®

Managing Tableau Bridge at Scale

By Miranda Osterheld, Senior Product Consultant

Contents

How to use this document

The first two sections of this document review fundamental concepts necessary for understanding why there are two main models for a Tableau Bridge deployment. If you are comfortable with Bridge, how it works, and its technical requirements, feel free to skip ahead to Different Models for Managing Tableau Bridge.

The purpose of this document is to inform customers on best practices for using Tableau Bridge Version 2019.2. Note that recommendations may change as new functionality is added in future releases.

What is Tableau Bridge?

Tableau Bridge is a free, supported client you can install behind your firewall to use in conjunction with Tableau Online, enabling connectivity to on-premises data. Tableau Bridge is currently only available for Windows operating systems. Previous versions of Bridge were a part of Tableau Desktop installations but starting with version 2018.2, Bridge is a stand-alone client. Only one Bridge client can be installed per Windows user context on a single machine, but customers can associate many Tableau Bridge clients with their Tableau Online site to support their data connectivity needs.

Tableau Bridge can be used in the following ways:
· Maintain live connections to on-premises data
· Maintain published extracts of data sources that Tableau Online cannot reach directly

To understand which data sources require Tableau Bridge, please reference Tableau Online: Keep Data Fresh. Exceptions are worth noting, especially scenarios called out as not supported by Bridge. The general guideline is that if it is a cloud-hosted data source, Bridge is generally not needed, and the recommended path for ensuring connectivity is IP whitelisting.

It is worth noting, however, that there is nuance between what may be considered an "on-premises" data source and one hosted in the cloud but is still behind a firewall and not accessible via the Internet. Some data sources are **always** "cloud-native", such as Amazon Athena, Redshift, Azure SQL Database, Google Cloud SQL, etc., so Tableau Online will by default expect to connect directly via IP whitelisting when the native connector is used.

In the case where a data source isn't necessarily "cloud native", either Bridge or IP whitelisting can be used to access the data sources. For example, a Microsoft SQL Server or Postgres database could be on premise or hosted in Amazon RDS, and it is up for the user to decide which connectivity option is preferable.
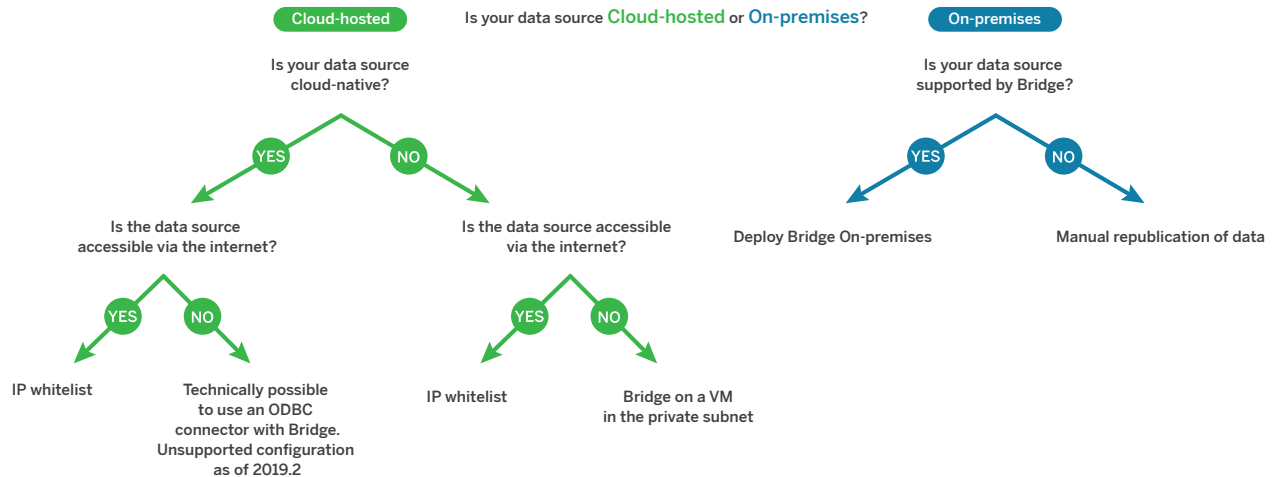
The diagram below provides more guidance:

Cloud-hosted
Is your data source
cloud-native?

Is your data source **Cloud-hosted** or **On-premises**?

On-premises
Is your data source
supported by Bridge?

YES   NO

Is the data source
accessible via the internet?

Is the data source accessible
via the internet?

YES   NO

Deploy Bridge On-premises   Manual republication of data

YES   NO

IP whitelist   Technically possible
to use an ODBC
connector with Bridge.
Unsupported configuration
as of 2019.2

YES   NO

IP whitelist   Bridge on a VM
in the private subnet

Figure 1 Connectivity Desicion Tree

It's technically possible to configure Tableau Bridge to work with "cloud-native" data sources if the data is completely isolated from the Internet in a private subnet (and therefore IP whitelisting is not an option). Bridge could be run on a virtual machine (VM) in a private cloud to make outbound connections, but a generic ODBC connection would be required to circumvent the direct connectivity Tableau Online expects. Consult with your Tableau representative to evaluate if this option is necessary and/or feasible.

Steps for installing Bridge can be found here. No significant hardware is needed to run Bridge clients; typically, small machines are acceptable (2 core 4 GB memory). Additional RAM may be useful for extract-heavy clients, and for Bridges running live connections, additional cores can be useful because live connections can be multithreaded.

How does Tableau Bridge work?

Tableau Bridge makes outbound connections only; all communication is initiated from behind the on-premises firewall, using ports 80 and 443. For live connections, Bridge uses WebSocket protocol to maintain a highly responsive connection to Tableau Online, in much the same way as some chat applications do. For extract refreshes, Bridge polls online to retrieve schedules and assigned data sources, and to publish extracts as scheduled. Data in transit, to and from Tableau Bridge, is encrypted. Bridge can only be used against published data sources, and it cannot be used against data sources embedded in a workbook.

There are two accounts that are relevant to keep in mind when conceptualizing Bridge:

1. The Tableau Online account used to log in to Tableau Bridge client and associate it with the user's Online site. Bridge does support SAML and OpenID to authenticate into Tableau Online if that is the associated authentication mechanism for that Online user.

Note: A user can only assign extract refreshes to Bridge clients that are running under their Tableau Online credentials. To assign an extract to a Bridge client, they also need to be the data source owner or a site administrator. They will only see clients they are authenticated into available for assignment.

2. The Windows account, because Bridge is executing all queries under the context of the local Windows user. The permissions associated with this Windows user are especially relevant when it is required that Bridge is run as Windows Service (see below), or when using Windows Authentication as the mechanism for authenticating into a published data source..

Bridge can run in two different modes: Application mode, or Service mode. By default, the client will run as an Application. This means the Windows user must be logged onto the computer where the client is running in order for scheduled refreshes to complete. Once signed in, the Bridge client will open from the system tray. If your client is set up to run in Service mode, you don't need to be logged on to the computer running the client, but your computer must be on. To enable Tableau Bridge to run as a service, the Windows user must be a local admin on that machine. These parameters often result in more advanced Tableau Bridge deployments using persistent virtual machines to run the Bridge, so that no individual publisher accidentally turns off their computer when it's set to refresh an extract overnight. More information on this is covered in the Different Models for Managing the Tableau Bridge section. When in Service mode, if a computer restarts, Bridge is designed to re-connect automatically—no need for an interactive sign in. However, site admins always retain the ability to delete Bridge clients from Tableau Online if they ever need to remotely disconnect a client, such as if a laptop running a Bridge client was stolen, for example.

Note: If you're using Tableau Bridge 2018.1 or earlier, even if live queries are enabled by the site administrator, it doesn't support live queries in Application mode. The client must be running in Service mode to support live queries. If you're using Tableau Bridge 2018.2 and later, after the site administrator has enabled live queries, live queries are supported in both Application and Service modes.

As mentioned, support for live queries to on-premises relational data sources must be configured at the site level by a site administrator. The site admin can then enable specific Bridge clients to support either "Extracts Only" or "Live and Extracts". After your administrator enables support for maintaining live connections, publishers will see this option during the publishing process. Site admins who log into Bridge clients will see these clients set to "Live and Extracts" by default, as opposed to the "Extracts Only" default setting for non-admin data source publishers.

Bridge does not support Initial SQL or Impersonation at this time. Bridge follows the same support guidelines as the rest of the Tableau platform: 30 months of support after a release (details here). It's anticipated customers will want to update fairly regularly to take advantage of continuous improvements.

Publishing data sources and scheduling extracts with Bridge

When a data source publisher goes through the publication process for either live connections or extracts, it's crucial to ensure Bridge has the credentials or access information it needs (such as a UNC file path) in order to perform the designated task—either live query or refreshing an extract.

Extracts

For extracts, access information (credentials/filepath) must be embedded in the Bridge client. The Bridge client owner must log in to the Windows machine and manually enter the credentials. This process results in database credentials being stored on the computer using Windows credentials manager. For data sources accessed via Windows Authentication, there are no credentials to embed, but the Windows account that Bridge is running under must have access to the source SQL server.

Tableau Bridge can perform one extract refresh at a time. You can schedule extracts to run at

different times to avoid conflicts, or rely on Bridge's default behavior, e.g. waiting until one extract has completed before starting the next. If you need to refresh more than one extract at the same time, consider setting up additional clients on different computers (or under different Windows user contexts) to perform the extract refreshes. More on this can be found in the Scaling Tableau Bridge section. If you choose to run multiple Bridges under different Windows users on a single machine, it's possible to assign extracts to each of those Bridge clients, and have those extracts run at the same time, but it's important to ensure the machine has the appropriate resources in order to support those simultaneous extracts. Many admins may find it preferable to simply utilize separate VMs running under a service account.

To ensure that long running refresh tasks don't prevent other extracts from refreshing, you can enforce a time limit (also known as a timeout limit), for refresh tasks performed by a Tableau Bridge client. The timeout limit is the longest allowable time for a single extract to complete a refresh task before it's canceled.

By default, the timeout limit for a client to perform a refresh task is set to 24 hours. If a refresh task exceeds the timeout limit set in the client, the client cancels the refresh task. If an extract refresh is canceled as a result of a timeout, you will be notified in the client and an email alert will be sent to the data source (in this case, extract) owner.

Note: To optimize performance on the server, refresh tasks are distributed within a range of up to five minutes of the time of day you specify. For example, if you set an hourly schedule, the occurrence set to run at 1:00 a.m. could run any time between 1:00 and 1:05 a.m.

As much as possible, it's recommended to schedule refreshes during times that don't conflict with business hours. Scheduling for less busy times (for both your business and generally, for Tableau Online) can have performance benefits.

If you work in a geographically distributed environment, your Bridge agents may be running in different time zones. When you schedule a refresh in Tableau Bridge, the time of day that appears there and on the Tableau Online schedule list reflects the time zone of the computer on which you are running Tableau Bridge.

If the computer running the Bridge is not on at the time when a scheduled refresh is set to occur, the extract will not occur, and the Bridge will attempt to take the extract again at the next scheduled time. The Bridge does not have any automatic failover capabilities when it comes to failed extracts.

Live connections

For live connections, if the data source requires authentication by the database, the credentials must be embedded in the data source at the time of publishing. If the credentials are not embedded in the data source at the time of publishing, the credentials can be added to the published data source within Tableau Online. This is because for live connections, the Bridge receives credentials from Online at the time of querying. The exception is when the data source is set up to use Windows Authentication, which means the query is executed by the windows account the Bridge is running under, so no credentials are embedded; the Windows account must be permissioned to access the data it needs on the SQL Server.

When more than one Tableau Bridge client supports data sources with live connections for a site, by default, all the clients are automatically included in a "pool." The purpose of this pool is to distribute (or load balance) the live queries among the available clients. Although the client in the pool that performs the live query work is chosen via round-robin, if for whatever reason a client can no longer perform the live query work another available client from the pool is automatically chosen to handle the live queries that are sent to it. There is no additional intervention required from the site administrator, or data source publishers, to support or manage the pool of clients.

Different models for managing Tableau Bridge

Generally, there are two deployment "extremes" that customers can use to set up Tableau Bridge: a Centrally Managed VM Bridge Deployment, and a Private Publisher Model. It is worth noting that many Bridge deployments may start as a Private Publisher deployment and evolve into a Centrally Managed deployment as more and more users, data sources, and Bridge clients are added. Hybrid deployments, where some Bridge clients are centrally managed, and others that are managed by individual publishers, are common.

Centrally managed Bridge deployment

For centrally-administered Tableau Bridge deployments, a site administrator will set up one or more Bridge clients to keep the data published by their data source publishers fresh.
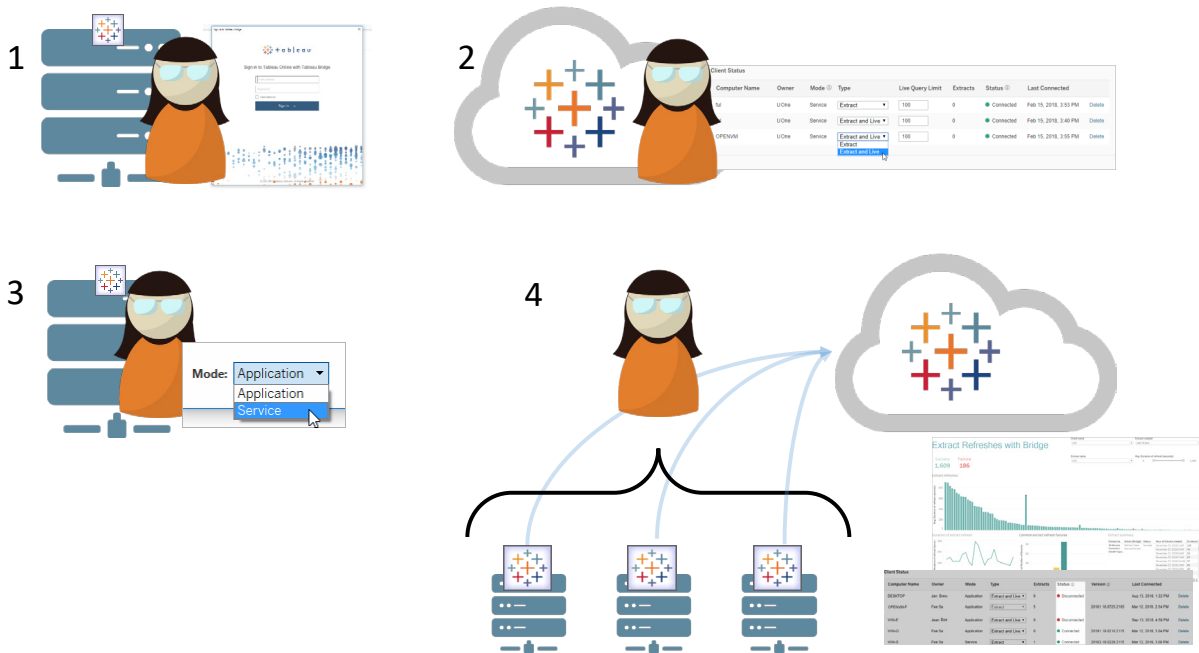

Figure 2 Site Admin Bridge set up steps for a centrally managed bridge deployment.

1. The site admin will install a Bridge client on a machine, usually a persistent Virtual Machine that is running under a Windows Service account. She will authenticate into the Bridge client using their Tableau Online credentials.

2. The site admin will ensure the connection to her Tableau Online site from the Bridge client is successful by logging into her Tableau Online site and checking that the client appears under "Settings" and "Bridge" tab. If relevant, she will enable that particular Bridge client for both live connections and extracts.

3. As relevant, within the machine Bridge is running on, the site admin will set up the Bridge client to run in Service mode to support live connections even when the Windows user is logged out of the VM.

4. The site admin will go through this process for each Bridge client they are setting up. She will monitor connected Bridge clients and data sources within the built-in views in Tableau Online.

Once Bridge clients are set up, a publisher would go through the following process to publish a data source using a centrally managed Bridge:
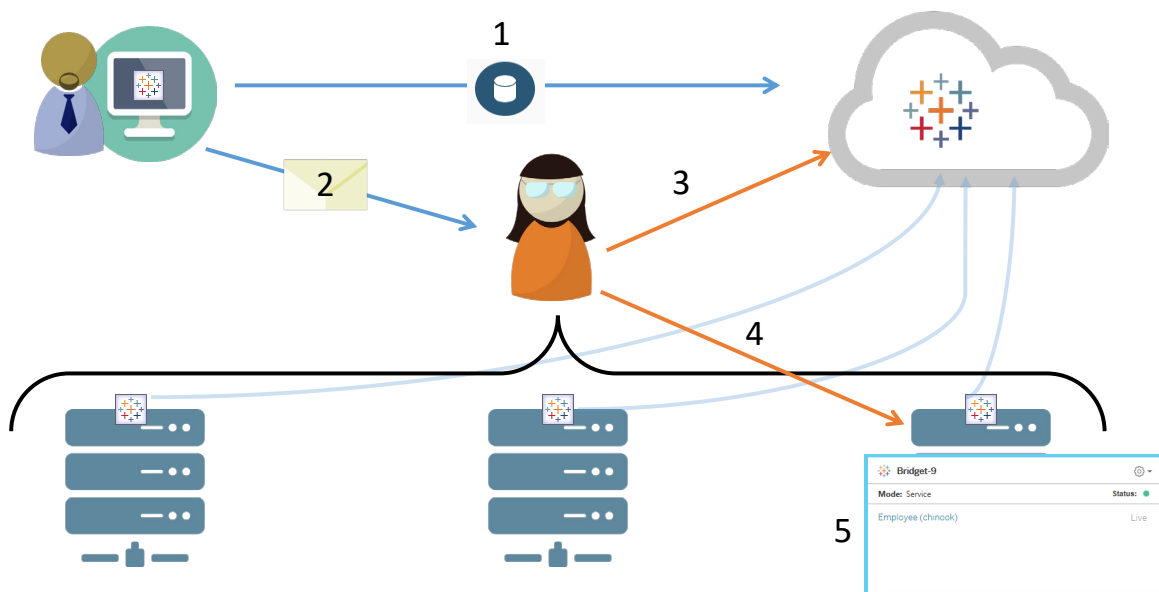
Figure 3 Workflow for a centrally managed Bridge deployment.

1. A data source publisher publishes a data source that requires the use of Bridge. Tableau Online will notify them that they need to assign the data source to a Bridge client.

2. The publisher will notify the site admin managing Bridge deployment that the data source is published, and provide relevant information such as the name of the data source, the schedule for refresh (if an extract), and any relevant access information (credentials, UNC file path, etc.). This communication happens outside of Tableau Online, usually via email.

3. The site admin takes the information and validates that the published data source exists in Tableau Online by reviewing published data sources in the "Explore" section. For live connections, the site admin embeds the credentials in the data source within Tableau Online. For extracts, the site admin assigns the data source to a particular Tableau Bridge client.

4. The site admin also validates that the data source has appeared in Bridge clients. For live connections, she can check any of the clients, as live queries can be sent to the entirety of a Bridge pool. For extracts, the site admin must enter the appropriate access information in the client the extract is assigned to (file path, credentials, etc.)

Notably, this model requires all Bridge clients be signed into Tableau Online under the site administrator's account so that the admin can assign work to those Bridge clients. Users are not able to assign work to Bridge clients they aren't authenticated into with their Tableau Online credentials. Site admins are not able to assign extracts to or manage extracts that other site admins own. Many customers use several small persistent virtual machines to run these Bridge clients and will use a dedicated service account for these machines.

Note: If you're using the same central client to handle both scheduled extract refreshes and load balancing of live queries, performance of the client could be affected if there are a significant number of live queries processing at the same time. To optimize for performance, it's recommended to designate some Bridge clients as extract only and designate other clients to handle the load balancing of live queries.
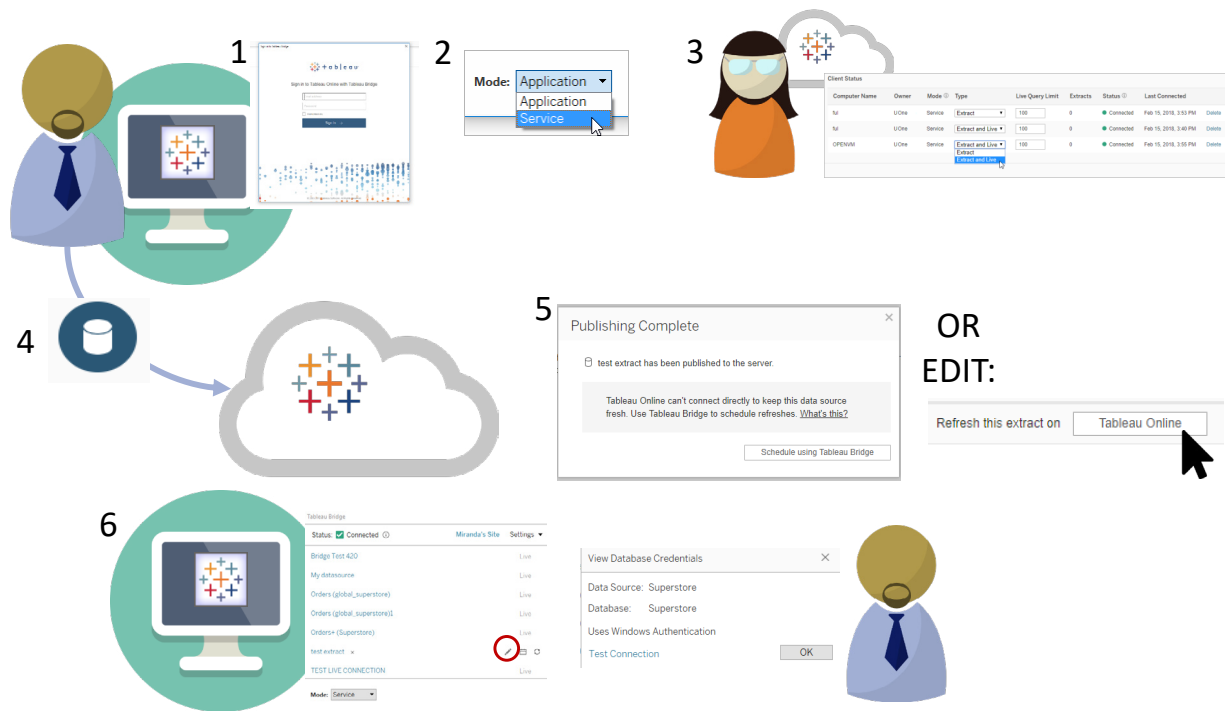
Private publisher model

Figure 3 "Private Publisher" Model Workflow

Bridge set-up steps by the private publisher

1. A data source publisher starts a Bridge client on their laptop or computer. This is either under Server > Start Bridge Client (version 2018.1 and prior), or Server > Install Tableau Bridge Client (for v2018.2+, when Bridge became a standalone client). The publisher needs to log into Tableau Online through the Bridge client with their Tableau Online credentials.

2. Optional, but recommended: The publisher should switch Tableau Bridge to Service mode on their machine in order for that Bridge to support live connections and extract refreshes even when the Windows account is logged out.

3. Optional, but often necessary: A site admin should ensure the bridge client is appearing under the "Bridge" tab of Settings in Tableau Online. This is also where they can ensure a Bridge client is set up for live connections.

Data source publication (extracts)

4. The data source publisher publishes the data source to Tableau Online.

5. When the successful publication notification appears in Online, the publisher has the option to schedule a refresh for the extract in that moment, or they can go to the extract later and select the button to assign the extract to a particular Bridge client.

6. The data source publisher returns to the Bridge client and ensures the data source has appeared. He enters the UNC file path or credentials into the Bridge client by clicking the pencil icon.

Data source publication (live)

4. The data source publisher publishes the data source to Tableau Online. For live connections, the publication dialogue will allow the publisher to select if they want to maintain a live connection using Bridge, or take an extract that can later be refreshed.

5. When the successful publication notification appears in Online, the publisher should ensure that the credentials necessary to access the data source have been embedded in the connection within online. Bridge receives those credentials at that time the query is sent from online.

6. The data source publishers ensure that the published data source appears in the Bridge client.

Scaling Tableau Bridge client

In many Bridge deployments, more than one Bridge client will be necessary to support data freshness needs. To decide how many extracts may be necessary, the limiting factor is generally how many extracts you can complete in series per day. However, it's likely that business requirements do not allow all extracts to be distributed in sequence throughout the day; more often, customers have several concentrated time blocks when they want their extracts to occur. Any two or more extracts that need to occur at the exact same time (or the extract completion windows overlap) require separate bridges per extract.

For Tableau Bridge clients supporting live connections, the driving factor is number of concurrent users. Site admins can monitor traffic to data sources with live connections using in a built-in administrative view in Tableau Online. This should give a high-level view into how often particular data sources with live connections are being accessed.

As a benchmark, our present internal Tableau Online development environment contains over 1500 users, with 8 clients supporting 160 unique published data sources configured for live connections. After monitoring extract times and failure rates using the provided administrative views, these clients were added until 8 clients provided acceptable performance. Those Bridge clients run on commodity hardware that was available at the time of configuration. Adding Bridge clients is simple. The client is simply installed and authenticated by the person who will be managing that client, depending on which model referenced above is being used. It's worth noting that if you're transitioning data sources to a centrally managed client, the site admin will need to switch the extract within Online to assign it to a Bridge client they own, and then ensure they enter credentials or the file path to that data source into the new Bridge client. Ideally, the site admin will have their own credentials for the underlying data source they want to embed, as opposed to receiving and embedding the original data source credentials into the new client. Since live connections receive credentials at the time of query, there is no transition needed.

Something to note is that if the only aspect that is changed is the Bridge client that the published data source is assigned to, the data source owner will still get notified for any failures, as opposed to the "owner" of the Bridge client. It's possible to change the owner of the data source to the site admin managing the centrally managed Bridge deployment, but the owners of the workbooks need to have connection privileges to the published data source to prevent connections from breaking when the site admin takes over the published data source. When site admins take over live connections, it's crucial they re-embed credentials into the published data source within the Tableau Online interface, because previously embedded credentials are cleared when data sources change ownership.

Monitoring Tableau Bridge client

As a site administrator, you can monitor Tableau Bridge client in a couple of ways:

· Monitor traffic to data sources with live connections using the built-in administrative view
· Monitor client status on the site settings page
· Monitor Bridge extract performance using the built-in administrative view
· Set up data source owners for extract refresh notifications

It's important to note that data source owners are the users who receive emails when data sources fail. This is especially relevant in the centrally managed model, because the owners of the Bridge clients are the site admins, and it is commonly necessary to get them involved when issues occur and need troubleshooting. Because they will not be automatically notified, data source owners who are not the owner of the Bridge client generally set up automatic email forwarding to site admins to keep them up to date on issues.

As a data source publisher, if you believe an extract is taking an unusual amount of time, there are several things you can do to troubleshoot. An extract refresh through Bridge has two components: taking the extracts locally with Bridge, and then sending that extract to Online and having it become available. To isolate the potential cause of the issue, it may be prudent to take the same extract in Tableau Desktop and compare the extract duration with the time it takes for Bridge to take an extract. Any anomalies there should be reported to your Tableau contact for review and support.

Summary

Tableau Bridge is a necessary component to many successful organizations leveraging Tableau Online and looking to ensure connectivity for their "on-premises" data sources. Bridge can be effectively scaled to support both extracts and live connections in order to meet these needs, and there are multiple ways to manage a Bridge deployment, depending on the preferences of an organization.