# HCLTech | Supercharging Progress™

# Scaling and deployment of Tableau Bridge

# Introduction

Tableau Bridge client is a free, supported software that can be installed behind a firewall to work with Tableau Online, enabling access to on-prem data. It is available exclusively for Windows operating systems. Each Windows user can install only one Bridge client on a single machine. However, customers can link multiple Tableau Bridge clients to their Tableau Online site to support their data connectivity requirements. Tableau Bridge can be used in the following ways:

- Maintain live connections to on-prem data

- Maintain published extracts of data sources that Tableau Online cannot directly access

For cloud-hosted data sources, a Bridge is unnecessary but IP whitelisting is the recommended connectivity method.
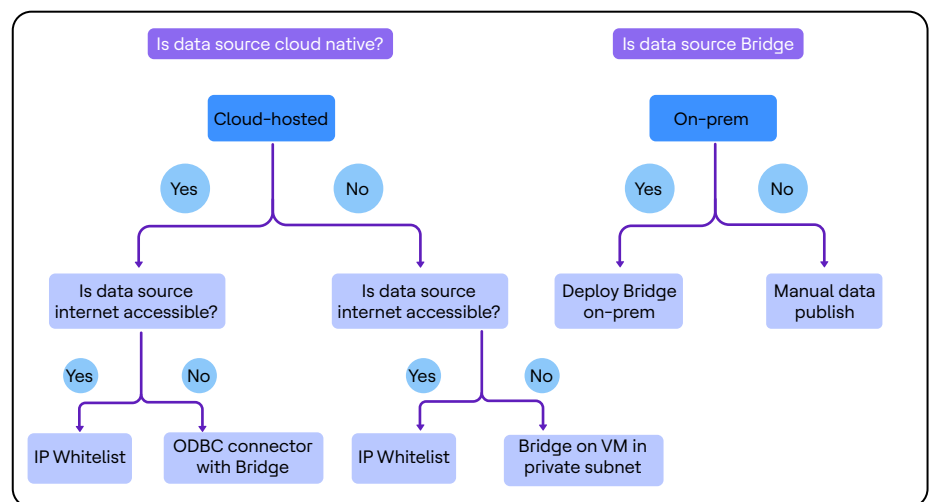


Tableau Bridge can be configured with cloud-native data sources if the data is entirely isolated from the internet in a private subnet. The bridge can operate on a virtual machine (VM) in a private cloud to establish outbound connections. However, a generic ODBC connection is beneficial for direct connectivity to Tableau Online. Running Bridge clients does not require significant hardware, small machines with two core 4 GB memory can be provisioned. For extract-intensive tasks, additional RAM can be advantageous, and for Bridges running live connections, additional CPU cores are beneficial since live connections can utilize multithreading.

# Functioning of Tableau Bridge

Tableau Bridge establishes outbound connections using only ports 80 and 443. For live connections, it employs the WebSocket protocol to connect to Tableau Online. For extracts, Bridge retrieves schedules and source information from the Tableau server to publish extracts as scheduled. Data in transit is encrypted. Bridge can only be used against published data sources, not embedded workbook data sources. A limitation is that any changes to custom SQL in the published data source must be implemented and deployed during release cycles, which can be time-consuming.

- Bridge supports SAML and OpenID for authentication into Tableau Online, provided the user has a valid Tableau online account

- Since Bridge executes all queries from local Windows users, appropriate Windows account permissions should be applied

Bridge operates in two modes: Application mode and Service mode. By default, it runs as an application, meaning that the Windows user must be logged in for a scheduled refresh to complete. For Tableau Bridge to run as a service, the Windows user must be a local admin on that machine. When deploying Bridge on VMs, these parameters are essential to ensure the server is not turned off during an overnight refresh. In Service mode, if the system restarts, the Bridge will automatically reconnect. Site admins can remotely delete Bridge clients from Tableau Online if necessary. A site administrator must configure live queries to on-prem relational data sources at the site level. The site admin can enable specific Bridge clients to support either 'Extracts Only' or 'Live and Extracts.' Once the connection support is established by the administrator, publishers will see this option during the publishing process.

# Publishing and scheduling using Bridge

When a data source is published via live connections or extracts, it is crucial for Tableau Bridge to have the necessary credentials to perform the required task, whether it is executing a live query or refreshing an extract. For extracts, the access information (credentials or UNC file path) must be embedded in the Bridge client. The Bridge client owner needs to log in to Windows and manually enter credentials in the Windows credentials manager. For data sources using Windows Authentication under which Bridge is running must have access to source SQL server. Extracts can be scheduled at various times to avoid conflicts and  Bridge's default behavior is to wait until one extract is completed before starting the next. To refresh multiple extracts simultaneously, additional clients must be set up on different computers under different Windows user contexts.

A timeout limit can be configured to ensure long-running refresh tasks do not prevent other extracts from refreshing. The limit is the maximum allowable time for a single extract refresh task before it is canceled. The default timeout limit is set to 24 hours. If a refresh task exceeds this limit, the client will cancel the task. If an extract refresh is canceled due to a timeout, an email alert is sent to the data source or extract owner. It is best practice to schedule refreshes during off-peak and non-business hours for optimal performance and system availability. In geographically distributed environments, Bridge agents may run in different time zones. The scheduled refresh time displayed in Tableau Bridge and on Tableau online reflects the time zone of the computer running Tableau Bridge. If the computer is off when a refresh is scheduled, the extract will not occur, and the Bridge will attempt to take the extract at the next scheduled time. Note that Bridge does not have automatic failover capabilities for failed extracts.

In live mode, if the data source requires database authentication, credentials must be embedded in the data source at publishing. If credentials are not embedded during publishing, they can be added later to the published data source within Tableau Online. For multiple Tableau Bridge clients supporting data sources with live connections on a site, all clients are automatically included in a pool, which load balances the live queries among the available clients. The client performing the live query is selected via round-robin. If a client cannot perform the live query, another available client from the pool is automatically chosen.

# Deployment strategy

There are two deployment modes for Tableau Bridge: a centrally managed VM Bridge deployment and a Private Publisher Model. Hybrid deployments are also common, in which some Bridge clients are centrally managed and others are managed by individual publishers.

# Centrally managed Bridge deployment

For centrally managed 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.
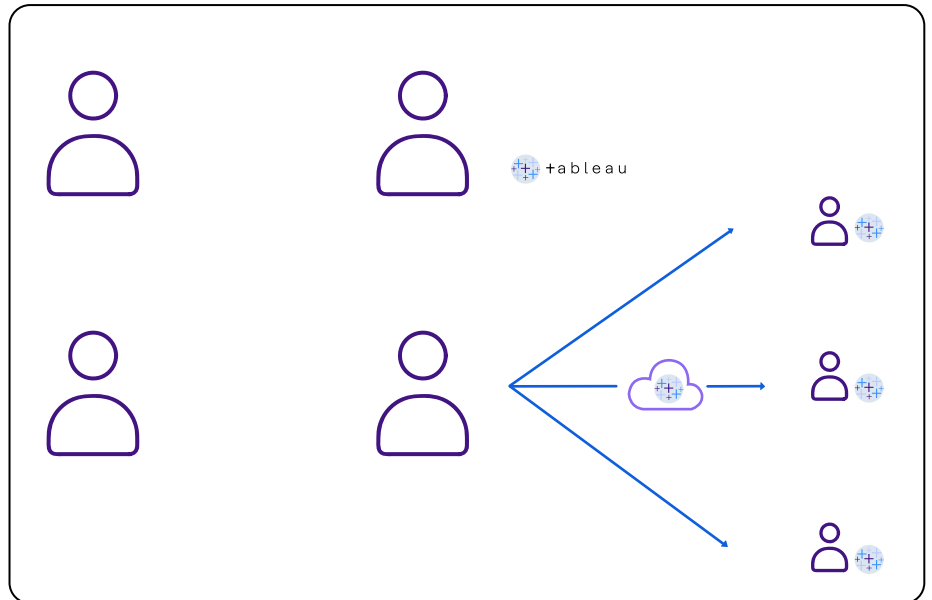
- The site admin will install a Bridge client on a machine. Usually, a persistent VM that is running under a Windows Service account will authenticate into the Bridge client through their Tableau Online credentials.

- The site admin validates the connection to Tableau Online from the Bridge client and ensures that the client appears under the Settings and Bridge tab.

- For the machine ON condition, 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.

- The site admin 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.
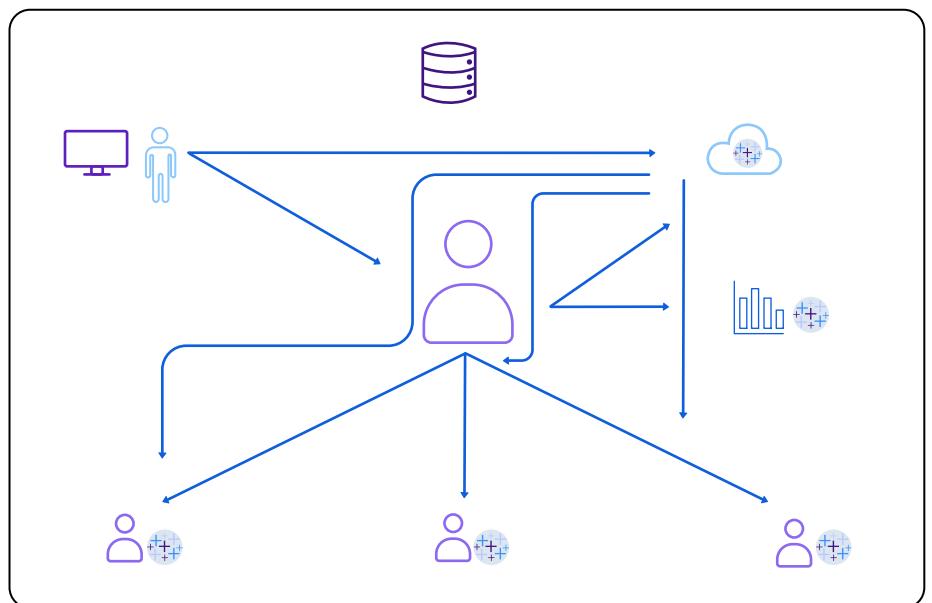
- Once a data source is published, Tableau Online will alert the publisher to assign it to a Bridge client.

- The publisher then informs the site admin responsible for Bridge deployment about the published data source, including its name, refresh schedule and frequency (if it's an extract) and any relevant access information (credentials, UNC file path).

- The site admin validates the published data source in Tableau Online by reviewing it in the Explore section. For live connections, the site admin embeds the credentials within Tableau Online. For extracts, they assign the data source to a particular Tableau Bridge client.

- The site admin also validates that the data source appears in Bridge clients.

# Private publisher model

- A data source publisher starts a Bridge client on their laptop. This is under either **Server> Start Bridge Client or Server > Install Tableau Bridge Client**. The publisher needs to log into Tableau Online through the Bridge client with their Tableau Online credentials.

- The publisher should switch Tableau Bridge to Service mode on their machine for that Bridge to support live connections and extract refresh even when the Windows account is logged out.

- A site admin checks that the Bridge client appears under the Bridge tab of Settings in Tableau Online. This is also where they can ensure that a Bridge client is set up for live connections.

# Scaling Tableau Bridge client

During deployment, more than one Bridge client is required to perform data refresh activity. The number of data extracts depends on the amount of data refreshes completed in series per day. However, the business does not allow all extracts to be distributed in sequence throughout the day as the users prefer specific time blocks for 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. The primary consideration for Tableau Bridge clients supporting live connections is the number of concurrent users. Site admins can monitor live connection traffic using a built-in administrative view in Tableau Online, which provides an overview of data source usage. When transitioning data sources to a centrally managed client, the site admin reassigns the extract to a Bridge client they manage and enters the necessary credentials or the file path. Live connections do not require transition since they receive credentials at the query time. If there are any changes to the Bridge client, the data source owner will be notified of any failures, instead of the Bridge client owner.

The site admin managing the centrally managed Bridge deployment can take ownership of the data source, but workbooks owners must have connection privileges to the published data source to avoid broken connections. When site admins manage live connections, it is essential to re-embed credentials into the published data source within Tableau Online because any previously embedded credentials are cleared when data sources change ownership.

# Monitoring Tableau Bridge client

Tableau Bridge clients can be managed by site admins through the following:

- Monitor traffic to data sources with live mode using inbuilt traffic monitoring views

- Monitor client health status on site level

- Monitor Bridge extract performance metrics, for example, time taken for extract completion, data records processed

- Configure extract refresh notifications for data source owners.

Data source owners receive emails when data sources fail. In a centrally managed model, the site admins who own the Bridge client should be notified regarding troubleshooting. Since they are not automatically notified, data source owners must forward alerts and failure emails to site admins. If an extract takes an unusually long time, several troubleshooting approaches exist. An extract refresh through Bridge involves creating an extract locally with Bridge and then publishing it to Tableau Online. To determine the root cause, it may be sensible to create the same extract in Tableau Desktop and compare its duration with the time taken by the Bridge.

# Conclusion

Tableau Bridge is crucial for leveraging Tableau Online and ensuring connectivity to on-prem data sources. It can effectively scale to support both extracts and live connections, facilitating load balancing of extract processing and enabling centrally managed deployment of data sources. In a real-world scenario, Bridge deployment is a best practice for multi-user, multi-tenant setups.

# Author and organization details

## Abhijit Das:

Abhijit has 13+ years of IT experience in the areas of data warehouse, ETL, BI Reporting and AWS. He has a keen interest in various tools and technology. He likes to explore new features in technology and apply it to real world problems.

# HCLTech | Supercharging Progress™

hcltech.com