Tableau and the Google Cloud Platform

Leverage the scale and flexibility of Google's Cloud Platform to power your Tableau Server deployment

Madeleine Corneli, Product Management Senior Manager

Contents

Introduction

This whitepaper is a technical guide for deploying Tableau on the Google Cloud. We will address deploying Tableau Server using Google Cloud infrastructure, implementing security and governance practices using cloud-first solutions, and connecting to data stored in the Google Cloud.

Tableau is an industry-leading business intelligence platform that combines powerful self-service analytics with a scalable and secure infrastructure. Tableau is built to fit seamlessly into your data infrastructure—connecting to data from over 80 different sources and leveraging the latest in cloud infrastructure for flexible hosting and scaling.

This whitepaper will touch on the key elements to consider when investing in a Tableau and Google cloud deployment. Consider this whitepaper a valuable tool to help guide you as you expand your cloud ecosystem.

Tableau Server on Google Cloud Platform

Tableau Server can be deployed in many configurations on all major public clouds. Your deployment will be unique to the needs of your organization. This section addresses patterns we guide customers to follow when deploying Tableau Server at scale for an enterprise on Google Cloud Platform (GCP).

Virtual machine sizing & performance guidelines

One of the early decisions in planning your Tableau deployment is machine sizing and selection. The size of your user base, how they will interact with Tableau content, and your data environment are all key considerations. We provide resources to help guide you through this decision process, including a whitepaper discussing details of Tableau Server sizing and performance.

When you deploy Tableau Server on Google you will leverage Google Compute Engine, a service for hosting on virtual machines available on Google Cloud infrastructure. Sizing your deployment on the cloud is similar to sizing an on-premises deployment. Google offers a variety of virtual machine (VM) options with a broad range of CPU, RAM, and Storage allocation.

Tableau Server deployments should always meet our minimum hardware recommendations. The cloud equivalent of CPUs are referred to as virtual CPUs (vCPUs) and follow a 2:1 ratio. Therefore, n2-standard-16 provides the equivalent of 8 CPUs (16 vCPUs) and 64 GiB memory. If you are looking to trial Tableau Server or want a baseline hardware recommendation to use for testing your usage patterns this is a great size to start with.

As your Tableau usage grows—either by adding users, content, data sources, or all of the above—you may want to adjust the underlying hardware to maintain optimal performance. Deploying on the cloud makes it easy to adjust your hardware to suit the needs of your organization. In addition to general purpose machines series (such as the N2D discussed above), Google also offers machines that are optimized for compute or memory. We refer to adding more compute power or memory to a given machine where Tableau Server is deployed as "scaling up" (vertically scaling) your Tableau Server. "Scaling out" (horizontally scaling) refers to adding additional machines to your Tableau deployment, and is discussed in more detail in the high availability strategy section.

* Please note that cloud providers periodically release new machine types and generations. If these types are not specifically addressed in this paper, they may still be a good option for Tableau Server deployment. Be sure to follow our published hardware recommendations and test your deployment before moving to full production.

Choosing how to scale up—adding more RAM, vCPU or storage—should be determined by the specifics of your organization's usage patterns. In general, VizQL (the Tableau Server process responsible for rendering visualizations) response times scale with CPU and the Hyper data engine (the Tableau Server process that stores and queries data in-memory) performance scales with RAM. Understanding the pattern of your users (more reliant on extracts, live connections, etc.) will inform the best allocation of resources. Tableau offers tools to help you test and understand the performance of your Tableau Server. We recommend you use these tools and follow the testing methodology outlined in this whitepaper to obtain the most accurate results.

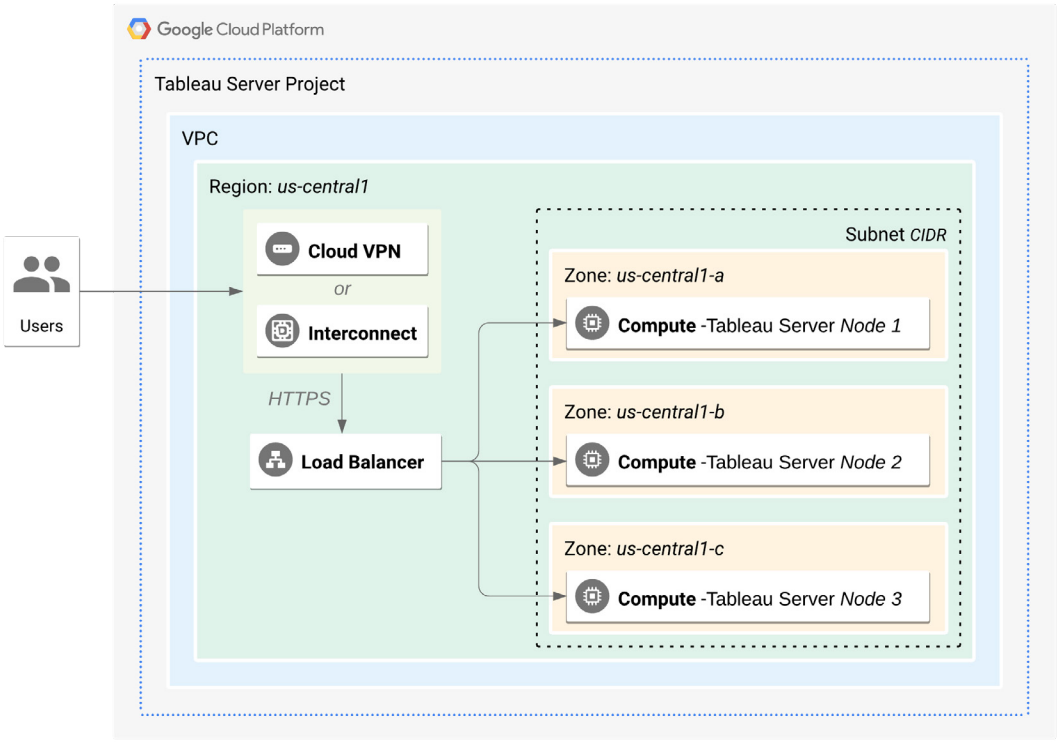
Distributed & multi-node Tableau Server deployments

Tableau Server can be configured to use multiple virtual machines, or "nodes"—this offers a few key benefits that allow you to:

- Provide Tableau Server with additional hardware resources that can scale to meet the needs of larger organizations (referred to as scaling out)
- Configure Tableau Server with multiple instances of various processes, eliminating single points of failure (requirement for high availability)
- Distribute Tableau Server across geographic zones to eliminate single points of hardware failure (requirement for high availability)

High availability

High availability (HA) is often one of the goals of a multi-node Tableau Server environment. By making specific choices when configuring Tableau Server and your Google environment you can create an infrastructure with redundancies preventing any single points of failure. Follow these instructions to understand requirements for Tableau Server to be highly available. In this whitepaper we will cover high-level recommendations for achieving HA—but you should also refer to Google documentation for guidance on designing robust systems.



This diagram provides a referenceable architecture for configuring a highly available, secure instance of Tableau Server on the Google cloud.

When designing a multi-node Tableau Server deployment we recommend distributing these resources to make your deployment more tolerant to outages. A multi-node Tableau Server instance can still run even if some of its nodes go down. Public cloud providers make resources available in different locations to support these exact types of resilient deployments.

* Please note that not all resources and machine types are available in every region. Make sure to consult the most recent GCP documentation while planning your deployment to ensure the resources you need are available in the locations you choose.

Google makes resources available in various different regions and zones. Zones refer to one or more data centers located within a single region. These zones are configured to balance traffic and resources across the hardware hosted within them. Regions are a collection of zones that are geographically grouped to decrease latency. You should deploy Tableau Server in a region (or regions) close to your users to minimize latency. You should deploy Tableau Server across multiple zones to help tolerate outages.

Deploying Tableau Server across three or more different zones can provide a suitable architecture for setting up a highly available, distributed environment. In the diagram above you can see these zones are us-central1-a, us-central1-b, and us-central1-c. You can refer to this documentation for instructions on how to set up a multi-node Tableau Server instance. For ease of management we recommend deploying all of your GCE resources for Tableau Server in the same project.

Please note that deploying Tableau Server across multiple regions offers additional tolerance against outages, especially multi-zonal outages (potentially caused by natural disasters or widespread outages). However, this type of architecture will require additional networking services and configuration.

Load balancing

When you deploy Tableau Server on the cloud we recommend implementing a load balancer to distribute incoming traffic to your Tableau Server. Google offers various types of load balancers with guidance on how to choose the appropriate one to meet your business needs.

In most cases, Tableau Server should be configured with load balancing at the 'application' level —meaning HTTP/HTTPS load balancing. These load balancers support SSL which you should configure for additional security for your Tableau Server deployment. If you are serving content only to your organization then you should consider an internal load balancer; if the content needs to be reached externally, or via the public internet, consider using an external load balancer.

The previous section offers details on deploying Tableau Server across multiple GCE VMs (often referred to as nodes). Next, you need to add each VM to an unmanaged instance group before they can be added to the back end of a load balancer. A similar result can also be accomplished using network endpoint groups.

*

Please note that Google offers three types of instance groups. Although managed instance groups offer additional functionality like autohealing, Tableau Server requires manual configuration that isn't easily compatible with these resources. You can leverage stateful managed instance groups by scripting adding Tableau Server nodes using instance templates if you need an automated solution.

Google provides step by step guides for setting up a load balancer within your Google Cloud Platform Virtual Private Cloud (VPC). Depending on your configuration you may want to use Google's DNS service or reserve a static IP through which users will access Tableau Server. You can refer to the diagram above to see how these services work together.

If you are setting up Tableau Server for internal use you should ensure that you lock down the entry point of your cloud environment to a Cloud VPN or a Dedicated Interconnect that is connected to your private network. For more information see the Google security whitepaper.

Once you have configured your load balancer in Google's Cloud console you should follow these instructions to add the load balancer to Tableau Server.

Securing network traffic

You should have a good understanding of Tableau and Google security practices before implementing a production instance of Tableau Server on GCP. Both Tableau and Google offer guidelines on network security best practices. Together, these two resources can help provide background on the various strategies and solutions available for building a secure environment. Tableau also offers a security hardening checklist.

VPCs & subnets

In Google Cloud you will manage all of your networking rules and access via a Virtual Private Cloud (VPC). Your Tableau Server deployment should have its own VPC, or should be added to an existing VPC you have configured. Fundamentally, VPCs are collections of subnets and their associated networking controls. Subnets are defined IP address ranges associated with specific regions. If you deploy Tableau Server in a single GCP region but across multiple zones you can assign a single subnet which makes managing access easier.

VPC's can be set to "auto mode" where, upon creation, a default subnet is added for each region. You can also create VPCs and subnets manually, which offers additional control. These options are discussed in Google's documentation.

Please note that some services need to be configured at the organization level and require higher admin level access than just project owner.

Accessing other Google resources

If you are deploying Tableau Server on GCP you may also be using other Google solutions, such as BigQuery. While both of these solutions will be hosted on Google infrastructure, it is common for enterprises to use separate projects and VPCs to separate and secure various elements of their cloud deployments. If your data lives on the Google Cloud (e.g. BigQuery) but in a different VPC than your Tableau Server instance then you can use VPC Service Controls to allow access between projects. VPCs can also be shared across GCP projects. Depending on your organization's requirements this may be an appropriate networking strategy.

Use firewalls to manage external access

In general, you should ensure that your VPC network firewall rules are configured to allow access on port 80 or 443 only, with the source limited to hosts or ranges of hosts that will access Tableau. For more information about firewall rules, see the Google page Firewall Rules Overview.

Examples of Tableau Server access that can be managed by firewall and networking rules:

- Connecting to data in other clouds
- Connecting to data on-premises
- Allowing users in separate networks to access Server (e.g. via VPN)
- Using connectors that rely on public APIs

Load balancing

When you configure Tableau Server on GCP with a load balancer, you will need to configure the load balancer to allow traffic from relevant sources. Google provides documentation on how to configure security policies to apply to HTTP(s) load balancers, like the ones we have discussed earlier in the paper.

HIPAA-eligible deployments

Your business requirements will determine the ultimate architecture you design for your Tableau Server deployment on GCP. Many industries have specific requirements that dictate security and data governance. The healthcare industry, for example, has strict guidelines for securing personal health data. For this reason, we have published a companion whitepaper that addresses the specifics of deploying Tableau Server on GCP in a HIPAA eligible architecture. Many of the principles outlined in these two whitepapers can be used to build a secure, compliant deployment that also facilitates self-service analytics at scale for your business.

Cost management

The Google Cloud Platform offers cloud-based services on a pay-as-you-go basis. Costs are determined by the services you run and amount of time you use them. Different combinations of instance types and sizes have different costs. In order to make it easier to manage these various costs and control them as you grow your cloud deployment we have created a series of Tableau dashboards for monitoring your GCP costs. If you are leveraging any GCP resources alongside Tableau we recommend using these dashboards for increased cost visibility.

Google provides a detailed overview of the costs of various services. You can estimate your total monthly costs using the Google Cloud Platform pricing calculator. You can also compare on-premises vs the cloud using the Google Cloud Platform TCO calculator.

To help monitor and control usage costs on an ongoing basis, you can set up billing alerts for the Google Cloud Platform to alert you when your monthly Google Cloud Platform costs reach your predefined spending threshold. For more information, see Set budgets and alerts at the Google website.

Identity & access management

Tableau Server has multiple layers of customizable security which can integrate with your organization's identity and access management solution. Authentication verifies a user's identity. Once authenticated a user is authorized to access and interact with certain content in specific ways. Finally, you can implement data security to limit users' access to underlying data.

Storing user identities

In order for individuals to authenticate into Tableau Server they must first be represented as a user in Tableau Server's identity store. While initializing Tableau Server you must select the source of the identity of your users. This choice can also inform the method of authentication. Your selection will depend on the requirements of your organization, the details of your deployment environment and any existing services or protocols you intend to use.

Tableau can be configured with a local identity store or an external identity store using Active Directory (AD) or an arbitrary Lightweight Directory Access Protocol (LDAP) server. In the local identity case, all identities are created and maintained within the Tableau Server repository. For external identity stores, users' identities are copied within the Tableau Server repository in the form of system users, but the external directory is the source of truth about their identity. Both scenarios can be configured to support single sign-on using various protocols.

External identity stores that support the LDAP protocol (such as AD or an arbitrary LDAP directory) support automatic group sync in Tableau. For all Tableau deployments (regardless of identity store type) you can also always manually add users (individually, or via CSV) or sync them via a script utilizing the REST API.

* Please note that when configuring Tableau Server, you cannot change the identity store once you have completed installation. If you need to change you will have to uninstall and reinstall Tableau Server. For details, refer to our authentication documentation.

Authenticating users with Google

You can configure Tableau Server to use Google as an IdP to authenticate users via OpenID Connect (OIDC) protocol—an overview is provided in our documentation. This configuration has been tested extensively and is popular with our customers as it is straightforward to implement and manage. If you choose this route you must configure Tableau Server with a local identity store. Our documentation covers the requirements for using OIDC with Google and Tableau. We provide detailed guidance on configuring Tableau Server for OIDC as well as configuring your OIDC IdP for Tableau Server, both of which will be necessary for this implementation.

Alternatively, you can configure Tableau to use Google as an IdP to authenticate users via SAML. If you would like to use AD or an LDAP identity store instead of a local identity store then SAML is available as an authentication method. We provide documentation for how to set up SAML in various configurations for Tableau Server.

Tableau Server also supports various other authentication methods including Kerberos, SSPI, Trusted Auth and Mutual SSL as well as popular IdPs and auth providers such as Okta and OneLogin.

Active Directory, Azure Active Directory, & Tableau

Tableau offers various options for integration with Active Directory (AD) and Azure Active Directory (AAD) for user management and authentication. AD and AAD are industry standards for identity management and can be used with numerous services and public clouds.

Google offers options for both federating Google identities with Active Directory and federating Google identities with Azure Active Directory. As more and more enterprises embrace multi-cloud models, cloud providers have invested in functionality that allows you to use services from various clouds together. For instance, you may choose to pair GCP, your cloud infrastructure provider, with Azure or Microsoft, your identity management service.

For additional information on how to deploy Tableau Server with AD or AAD as your identity service please refer to our Tableau and Azure whitepaper.

Data source authentication & access

Data source authentication in the context of Tableau Server occurs when a user interacts with a data source hosted on Tableau Server (by viewing a dashboard, creating a new workbook, connecting to a hosted data source from Tableau Desktop, etc.) or when Tableau Server interacts with a data source on behalf of its users (refreshing an extract).

Data source authentication may be independent from Tableau Server authentication and by default Tableau Server does not act as a proxy to data sources. Just because a user has access to Tableau Server does not mean they are guaranteed access to data sources. You can learn more about best practices for data sources on Tableau Server.

Upon publishing a data source to Tableau Server, you have several options—you can prompt users to enter their own credentials to the data source, embed credentials, or enable an SSO experience that passes user credentials back to the data source. When planning your authentication strategy, especially as it corresponds to data sources, your goal should be to streamline your end user experience (to keep them in their flow) while maintaining appropriate security. You can read detailed explanations about the various options for determining access to published data sources.

When publishing a production data source that multiple users will access, it often makes sense to embed the credentials of a service account from the source database. This account should permit all actions required by Tableau users interacting with the data, but should limit their capabilities for security purposes. When users connect to this Tableau data source they will automatically be authenticated to the backend database under the service account credentials.

Tableau & your Google data ecosystem

Tableau offers an extensive suite of tuned connectors that allow you to query data stored in the Google cloud. Tableau connectors support both live and extract-based queries which leverage either the computing power of your Google data solutions or Tableau's powerful in-memory engine, respectively. The breadth and flexibility of Tableau connectors make it an ideal analytics platform for the Google Cloud. Many Tableau customers rely on multiple data solutions and our 80+ native connectors allow you to analyze data from anywhere, on-premises or in the cloud.

Google Cloud data connectors

Google BigQuery is Google's serverless data warehouse that offers scalable compute and storage capabilities designed for business agility. With Tableau's BigQuery connector, customers can analyze billions of rows in seconds without writing a single line of code. Combining the cloud agility of Google BigQuery with the blazing speed of Tableau has helped organizations realize analytics project value fast, and stay connected through their data. Tableau's connector can leverage Google BigQuery BI Engine to supercharge live queries with no extra configuration. Tableau's connector can also connect to BigQuery ML to augment your analytics with machine-learning insights. We have published a comprehensive whitepaper that covers the BigQuery connector in more detail—including performance tuning and optimization guidelines.

Google Analytics is Google's Marketing Analytics platform. Many customers use the Tableau Google Analytics connector to analyze user trends and web advertising performance. Together, Tableau and Google Analytics can help you build a comprehensive marketing practice driven by self-service analytics. You can refer to our Tableau and Google Analytics whitepaper for more information on how these two platforms work together.

Tableau has connectors to both **Google Drive** and **Google Sheets**. These connectors make it easy to access and analyze your data in the cloud. Many people also use Google Sheets to power their Tableau Public visualizations!

Tableau also connects natively to **Google Ads**, and **Google Cloud SQL**. All together this comprehensive connectivity suite helps you build a complete data analytics experience with Tableau and Google.

Data transfer costs

Many public cloud providers, GCP included, offer data ingress for free, but typically charge for data egress. Google covers the various types of ingress and egress traffic and their associated pricing in their documentation. Google also offers tools for transferring data into BigQuery and Google Cloud Storage.

If you use a load balancer to distribute incoming traffic between your Tableau Server nodes, there is no charge for ingress—although there is a cost associated with deploying and maintaining the laid balancer itself. Traffic between VMs in different zones, which will occur if you follow our multi-node configuration guidelines above, are subject to small egress charges since data is leaving one VM to arrive at another—more details here.

Since ingress is not charged, querying data from both inside and outside of Google should not be a concern when deploying Tableau Server on GCP. In general, most Tableau customers do not find data transfer costs to be prohibitive to deploying Tableau Server on the public cloud.

Tableau data governance

Data is everywhere and so is the demand to access and analyze it. As our organizations generate more and more data, this process gets increasingly difficult. Effective, integrated data management is the key to bringing order to the chaos. Tableau Data Management combines various solutions to make managing your data easy and scalable. Tableau Prep Conductor allows customers to centralize the scheduling, monitoring, and administration of data preparation performed in Tableau Prep Builder. Tableau Catalog simplifies the discovery of trusted data and content for end users, while also giving IT additional capabilities to manage, monitor, and govern their environment at scale.

In addition to live data connections, Tableau Server also has its own optimized in-memory data engine. Hyper is designed for fast data ingestion and analytical query processing on large or complex data sets. Leveraging Hyper alongside your live data connections gives you the ability to optimize data performance and decrease analytical loads on your back end systems.

Part of a holistic data strategy is confidence in the security of the data you store within Tableau Server. Tableau Server offers encryption at rest which enables various compliance scenarios and improves the security of your Tableau Server deployment. For more in-depth information on data and content governance—including data source and metadata management, curation, content management, and more—read about governance in Tableau.

Multi-cloud & hybrid scenarios

According to Flexera's State of the Cloud report, 76% of enterprises said they're incorporating multiple public clouds to accomplish their business goals. More specifically, 49% of companies indicate that they are implementing apps siloed as part of their most common multi-cloud implementation. The cloud makes it easy to choose the right tool for the job. You may select Google as your infrastructure provider, Azure as your data provider and maintain an on-premises Teradata instance. Tableau's customers derive value from being able to deploy flexibly on the cloud and connect to data anywhere. As you embark on your cloud journey you should take confidence in the fact that Tableau will support your business wherever you go.

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 solutions for Google

Learn more about how Tableau integrates with the entire Google Cloud Platform, including native connections to Google Analytics, Google Adwords, Google BigQuery, and more.

Explore solutions →

Tableau for IT

Get valuable IT resources to enable your organization with modern BI, whether you're evaluating, using, or deploying and scaling Tableau.

Visit resource hub →

Tableau Blueprint

Working with thousands of customers and analytics experts, we've captured best practices and developed a step-by-step methodology to create a data-driven organization.

Learn more about Tableau Blueprint →

Special thanks

The author of this paper would like to extend a special thank you to the following contributors to this content:

Kevin Glover, Product Management Senior Manager, Tableau

Maruti C, Staff Technical Solutions Consultant, Google

Joel Walsh, Program Manager for Analytics Partnerships, Google