Next Generation Cloud BI: Tableau Server hosted on Amazon EC2
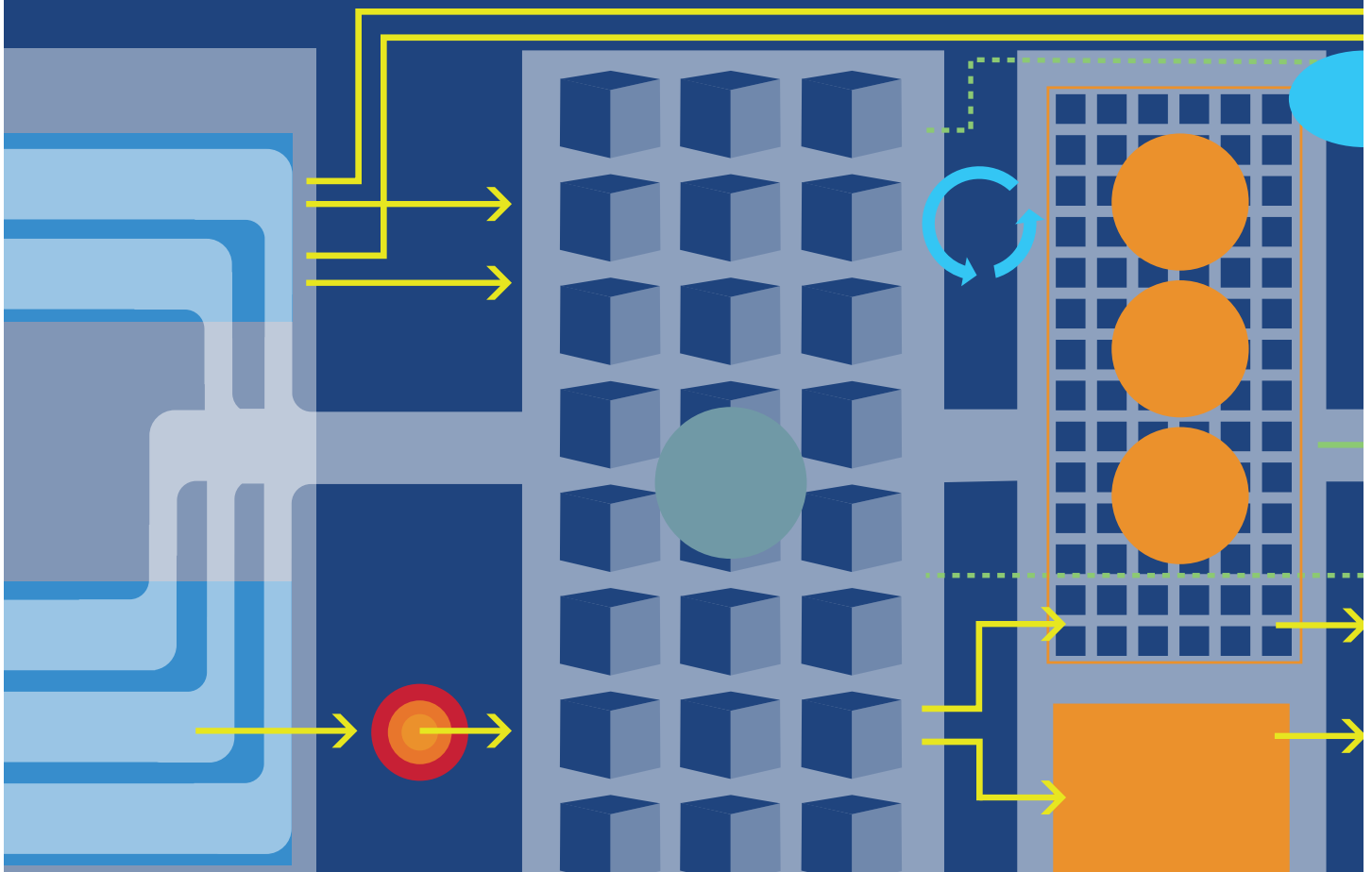
Table of contents

Introduction

Abstract

Amazon Web Services (AWS) and Tableau Software provide a comprehensive business intelligence solution that can be implemented quickly, secured easily, and used by everyone. Tableau enables organizations to analyze and share the data they have in Amazon Redshift, Amazon Elastic MapReduce (EMR), Amazon Aurora, Amazon Relational Database Service (Amazon RDS) and other AWS data services. This whitepaper outlines some of the reasons that enterprises should consider Tableau Server on Amazon Elastic Compute Cloud (EC2) to help see, understand, and share data so they can fully realize their investments with a mission-critical visual analytics platform on AWS.

The need for a modern business intelligence practice in the cloud

Digital transformation and a growing multitude of "things" are creating more data—and faster—than ever before. With increasing use cases for data and greater demand to access it for decision making, finding ways to make all of this data a useful resource for the entire organization is a complex challenge. However, with the right tools to empower people, data can be a tremendous asset. Organizations who drive insight across their increasingly diverse data streams give themselves a distinct competitive advantage.

Applications used for data analysis and visualization are gravitating toward the data itself. Today, that means a large-scale shift towards the cloud. Any organization can collect and present data, utilizing big data or database services from AWS such as Amazon Redshift, Amazon EMR, Amazon Aurora, or Amazon RDS. Whether an individual organization has an extensive, cloud-based big data practice or is currently doing very little analysis of their data, they can reap significant benefits by giving people across business and IT departments the ability to visualize patterns and analyze for insights it contains. Implementing a modern BI solution like Tableau can help them do that.

Amazon EC2 provides flexible, cost-effective, and high-performance virtual machine instances on the AWS Cloud, making it an ideal platform to host Tableau Server. With Amazon EC2, an organization can spin up as many virtual machine instances as needed to support their BI workloads, while circumventing upfront investments, hardware procurement, and deployment. Because activities like scaling and load balancing can be automated, Amazon EC2 also removes much of the manual effort that is required to enable high availability and optimal resource utilization on-premises.

When combined, Amazon EC2 and Tableau software can help organizations implement a modern BI practice quickly and get more out of their data. By drawing insights from the data produced by Amazon Redshift, Amazon EMR, Amazon Aurora, Amazon RDS, and a variety of other AWS services and sharing their findings across the company, organizations can fully realize their investments in AWS.

Deployment and scalability

The challenge to remain agile at scale

With legacy BI products in on-premises environments, deploying and scaling up is complex, time-consuming, and expensive. Every time an organization brings in a new data source, additional capital investments and lead time are required. Similarly, for the organization that wants to expand into a new geographic region, they must make additional investments, regardless of how many people they need to support.

Tableau is most well-known for an intuitive UI that can makes visual analytics accessible to business users across all skill levels. However, it's also flexible and straightforward for the IT professionals who are responsible for the underlying infrastructure and installing, maintaining, and updating the software. But to be able to take advantage of this scalability, modern BI requires agility at the hardware level. This can be difficult in many on-premises environments, but on Amazon EC2, many of these challenges can be circumvented.

The flexibility of the cloud

Amazon EC2 provides organizations with simple access to flexible, high-performance, cost-effective virtual machine instances, making it simpler to host and scale Tableau workloads. There are a variety of instance types available from AWS, automation for processes like scaling and load balancing, and pay-as-you-go pricing with no long-term commitment. Unlike on-premises deployments, Amazon EC2 allows an organization to spin up the capacity it needs in minutes, without hardware procurement or implementation. When that organization needs to scale up, they can do so instantly. Once Tableau is running on Amazon EC2, it can easily be integrated with many other AWS workloads, including Amazon Redshift, Amazon Aurora, Amazon Elastic MapReduce, Amazon Relational Database Service and other AWS products, plus third-party solutions.

The "Tableau on AWS" Platform
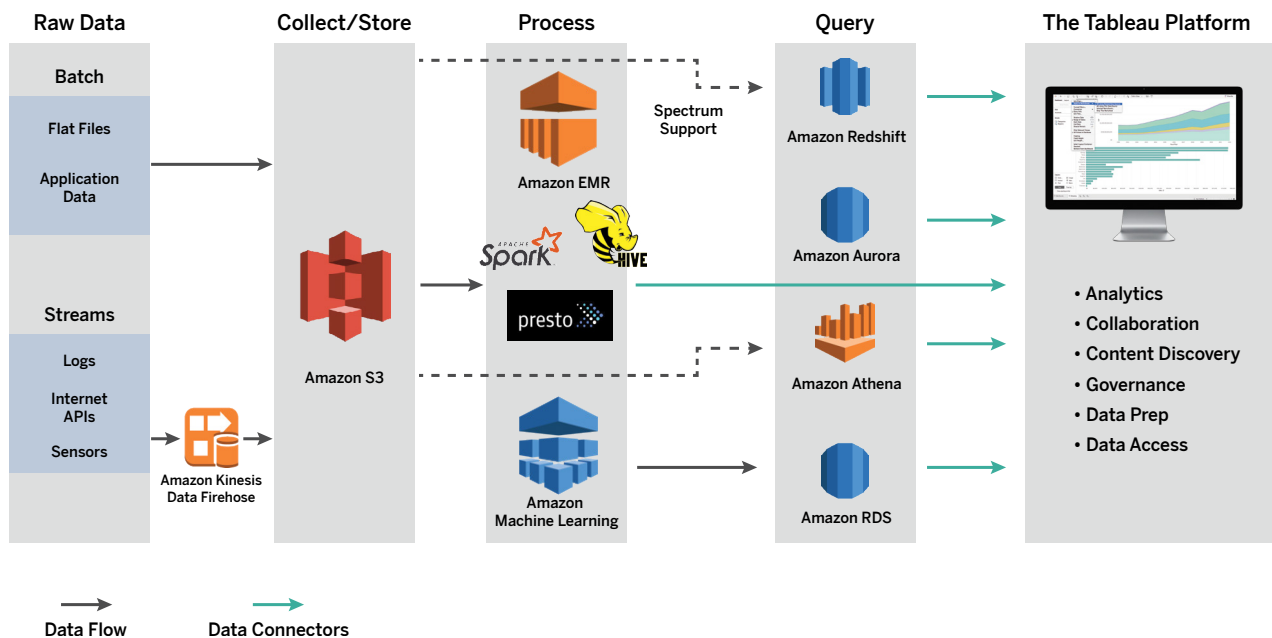
Security and governance

Data integrity and downstream analysis

BI practices can mean a lot of data—some or all of which is sensitive—is concentrated in a single place and shared across a large base. Many BI practices are user-friendly enough to gain widespread adoption within an organization, but lack the security required to protect an organization's data and intellectual property. Take, for example, the age-old practice of using spreadsheets for downstream analysis. Spreadsheets offer limited, if any, means for controlling access to the data. They can also be edited by anyone with whom they are shared, and there is no audit trail for admins to see who made what changes, when. If the formulas are broken or data is corrupted, there is virtually no way of determining where those errors occurred. Unfortunately, there are many more examples of the kinds of security and governance vulnerabilities that come with antiquated BI practices.

The security and governance you need

On top of the world-class security offered by Amazon in the cloud, Tableau Server provides a broad and deep set of security and governance features for projects and workbooks. To help enable strong data governance, Tableau Data Server allows administrators to define and certify data sources as well as standardize calculations and data fields for everyone to use. This helps minimize version control and security issues that can arise when data models are being shared across large, complex teams. As an added convenience, Tableau Data Server also removes the need to install database drivers on users' local machines.

Before publishing, administrators can specify who sees what data, and who sees which dashboards with filters and row-level security. For example, in a sales report that gets shared with regional managers, an organization may want to allow only the western regional manager to see the western sales, the eastern regional manager to see the eastern sales, and so on. With row-level security, you can give both managers the same view, but only allow each to see the data for their particular region. Integration with Microsoft Active Directory or other authentication processes makes it easy to set the permissions that determine what each person or group sees. For data sources that connect to live databases, administrators can also control whether people are prompted to provide database credentials when they click a published view. Tableau Server even has version control, making it easy to revert to a previous view if need be.

Dashboards published to Tableau Server can be securely accessed and edited from any browser, enabling the larger team to drill down and ask new questions by editing a view on the Web or on a tablet. Views can also be embedded into common web applications, like Salesforce and Microsoft SharePoint. In addition, Tableau Mobile allows anyone to access their data and dashboards from mobile devices, allowing them to author and view anywhere.
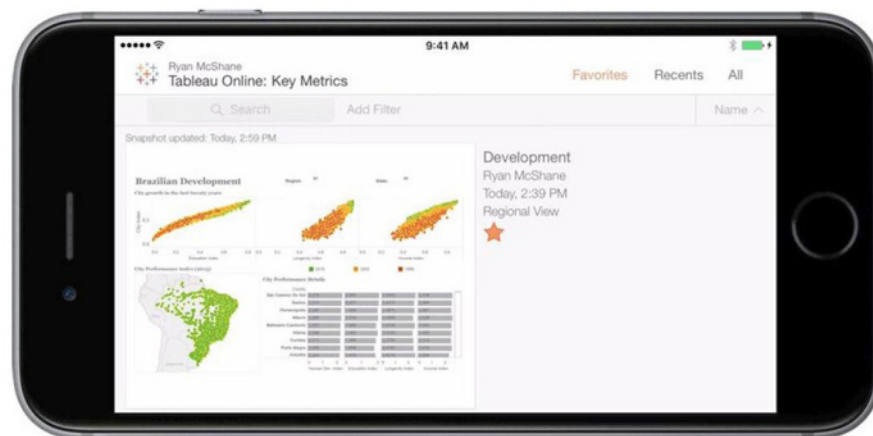
Self-service BI user experience
The user experience dilemma

One of the primary obstacles that organizations face in achieving these benefits is a low adoption rate of the BI products that they roll out. For someone whose primary role isn't related to data, most BI products are unnecessarily complex, offer a poor interface, or are not built to answer the questions they have, which discourages use. When individuals are given a BI product that is overly complex, they are likely to perceive that it isn't built for them, and they don't have any reason to use it. If a product has a cumbersome interface, they are likely to resort to their traditional ways of working, whether that be to act off of their gut, or to ask someone else within the organization.

Similarly, when people feel that a BI product they're given locks them into certain ways of slicing and visualizing the data that are not congruent with the types of findings they're looking for, they're unlikely to integrate that product into their processes. If a BI product cannot pass any of these tests with an individual, and that individual is not specifically required to use data gleaned from it to do their job, they will more than likely perceive that the potential benefits of that product are not worth the hassle of learning how to use it.

Breakthrough technology and faster visual analysis for everyone

Designed to support how people think, Tableau allows people to drag and drop multiple data sources onto a simple canvas, then view that data based on the variables of their choice and leverage their natural ability to spot visual patterns quickly. Anyone can ask new questions, spot trends, and identify opportunities, helping them make data-guided decisions. Tableau Desktop is compatible with simple spreadsheets, databases, complex big data environments, or virtually any other data source, on-premises or in the cloud. Dashboards authored in Tableau Desktop can be securely shared using Tableau Server, enabling people to share findings with partners or other members of the organization in a self-serve fashion and helping to broaden the application of data and data-driven decision making.

Patented technologies, including VizQL™, allow query and visualization to be combined into one process, so people can focus on their business problems and ask questions of their data. With Tableau Desktop, anyone can connect any of their data sources and produce visualizations that allow them to pose new questions, spot trends, identify opportunities, and make more data-guided decisions. It can also be used to build new calculations from existing data, drag in reference lines and forecasts, access statistical summaries and identify trends, regressions, and correlations. Tableau Server can be used to create interactive presentations where the presenter controls the narrative, and the audience is free to explore, which is useful for cross-team collaboration.

Additionally, the Tableau data engine, Hyper, uses proprietary dynamic code generation and cutting-edge parallelism techniques to help customers analyze large or complex data sets faster. This makes ad-hoc analysis on billions of rows of data possible with fast performance for extract creation and query execution.

Case studies

TripAdvisor grows per-user revenue with insights from Tableau on AWS

TripAdvisor, the world's #1 travel site, has been able to realize many of these benefits by leveraging Tableau on AWS. TripAdvisor's Revenue Optimization (RevOps) Department is responsible for increasing per-user revenue by 15% annually. To help reach this goal, they run roughly 45 A/B tests per fiscal quarter.

Before using Tableau, Product Managers (PMs) on the Revenue Optimization team would receive an email at the end of each A/B test, summarizing its findings. However, these emails would typically only answer half of the Product Manager's question, require manual math, and inevitably be lacking some critical information.

The RevOps PM would fire up a physical terminal and run a set of pre-built Hive queries, which would take some time, to hopefully get the detail they need to answer their question. Then, they would take that output to Excel, where they would spend several more hours using Pivot Tables, vlookups, and other calculations to hopefully arrive at the answer they need. If this query was still not giving them that answer, they would have to bother a software engineer to build them a custom query, which would tie up the majority of the day.

Eventually, TripAdvisor realized that this was not working for them, and that they needed a BI product that would allow them to: iterate faster as they generate new site ideas to generate additional revenue; query and gather the data needed to make decisions faster; increase confidence in their measurements; improve how results are communicated across the organization; and learn more from each A/B test. TripAdvisor also knew that they needed to find a way for daily reports to answer all of the PM's questions to help reduce the burden that they were placing on the engineers, and that the solution they rolled out needed to be easy to scale.

TripAdvisor decided to build a solution called TripDNA, which they deployed using Tableau Server, hosted on Amazon EC2, and pulling data from an Amazon Redshift data warehouse. They were able to deploy in under 3 hours, and have fundamentally changed the way that their PMs work. PMs can analyze how their tests are performing and answer questions that arise throughout their test with no need to pull in engineers.

As a result of using TripDNA with Tableau and AWS, TripAdvisor was able to increase per-user revenue by 22% over their first year, and 29% increase in the first half of their second year. Product managers have been able to do away with emailing cobbled together spreadsheets back and forth, which drastically reduces errors and manual calculation. Executives also have clear visibility into the team's progress against goals, higher confidence in the results being produced, and better collaboration with PMs.

How Netflix built its analytics in the cloud with Tableau and AWS

As Netflix has famously grown to support more than one-third of all internet traffic, the organization has needed to expand its data capabilities to suit. Its extensive analytics platform built on Tableau and Amazon Web Services offers a fantastic blueprint for organizations who are looking to build scalable and flexible business intelligence in the cloud.

With 86 million members and counting, Albert Wong, Netflix's Reporting Platform Manager, has to build everything related to analytics at scale. The company's data platform, shown above, is a complex but elegant system built on events and operational data fed into Amazon S3. Data is then sent, depending on end use, into Amazon Redshift or a series of NoSQL data processors that are then aggregated into Tableau Data Extracts. This data lake/data warehouse strategy allows Netflix to collect and store massive amounts of data, but also provide a high-level view of the data to analyze and explore. Of course, all of the data connections and extracts end up on Tableau Server, hosted on Amazon EC2.

Netflix uses Tableau Data Server so it can reuse its data sources and govern them across a wide range of users. For instance, one of the most important dashboards that Albert developed is one that shows usage and watch patterns within individual countries. With these dashboards, country managers can more easily manage the programming for their audiences.

Even though there are dozens of people that benefit from that dashboard, only one data source is used to feed it. Albert also sets permissions at both the group and individual level across the data source, ensuring that every team has access to the right information. By using Data Server, Albert can help his team answer multiple questions with a single governed dataset.

How to get started

You can self-deploy Tableau Server on an Amazon EC2 instance that you provision, deploy Tableau Server using the AWS CloudFormation templates in the Tableau Server on AWS Quick Start, or deploy Tableau Server on AWS using an AWS Marketplace Amazon Machine Image (AMI).

Tableau can be deployed on AWS infrastructure

Ways for the customer to get started

Tableau Server	AWS Self-Deployment	AWS Quick Start	AWS Marketplace AMI
Production ready	✓	✓	✗
Upgradable	✓	✓	✗
Scalable	✓	✓	✗
Distributed installation	✓	✓	✗
Active Directory support	✓	✗	✗
14-day trial license	✓	✓	✓
BYOL license	✓	✓	✓
Subscription license	✓	✓	✓
	Production		PoCs/Testing/Evals

The most flexible and scalable way to run Tableau on AWS is to spin up an Amazon EC2 instance, download Tableau Server, and install it on that instance. This will give your organization the most granular control and freedom to scale as business demands. For the best performance, try to stick to 16 vCPU instances, like the m4.4xlarge.

The AWS Marketplace offers one-click deployment of Tableau Server instances. The BYOL (Bring-Your-Own-License) instance types come with a free trial of Tableau Server, and if you choose to purchase you can obtain a license from Tableau. This allows for deployment of the Tableau Server AMI in an organization's AWS environment very quickly.

Tableau is also offered in 10, 25, 50, or 100 licenses per hour, and the price of those licenses is rolled directly into the AWS bill. This is the least flexible and scalable way to run Tableau on AWS and is most often used for a proof of concept. If license needs change, you must cancel and order again. In addition, clustering is unavailable with this option, which may not be ideal for especially complex or demanding workloads. However, this is by far the fastest and easiest way to get started running Tableau on AWS. There are also no long-term commitments, and everything is billed through AWS, which is very convenient.

About Tableau

Tableau is a complete, easy-to-use, enterprise-ready visual business intelligence platform that helps people see and understand data through rapid-fire, self-service analytics at scale. Tableau works with AWS services to empower enterprises to maximize the return on your organization's data and to leverage existing technology investments. Tableau Server runs seamlessly in Amazon's cloud infrastructure so the customers who prefer to deploy all of their applications on AWS have a complete solution offering from Tableau. Unleash the power of your most valuable assets: your data and your people.

Additional resources

Tableau Server on AWS Technical Deployment Guide

Tableau Server on AWS Quick Start

Tableau in the AWS Marketplace

Learn more about Tableau and Amazon Web Services

Learn more about Tableau products

Whitepaper: Optimize Tableau and Redshift Performance

+‡+ + t a b l e a u®