SIGMA

Cloud BI & Analytics with Snowflake and Sigma

Bring Snowflake Data to Life with
Cloud-Native Analytics from Sigma

Table of Contents

Abstract

For the better part of the last decade, data warehousing and analytics have remained stagnant. On-prem data warehouses and legacy analytics tools struggled to advance in meaningful ways, meanwhile, business intelligence (BI) adoption never took off like experts hoped. Conventional data warehouses and BI solutions failed on the promise to simplify data collection, enable rapid analytics, and generate actionable insights for business experts. But that has changed in recent years with the advent of the first data warehouse and analytics solution built entirely for the cloud.

Snowflake and Sigma present a revolutionary new approach to analytics. Snowflake's unique data architecture is designed to handle all aspects of data and analytics, resulting in a data warehouse that delivers the performance, flexibility, concurrency, and affordability not possible with other providers. Sigma builds on top of this model, leveraging the full power of the cloud to make data accessible for everyone. Together, Snowflake and Sigma have changed the cloud data conversation. Now, entire organizations can explore data, ask the tough questions, and share discoveries in real time.

This paper explores the common analytics challenges solved by Sigma and Snowflake, and shares the ways modern companies can put these revolutionary technologies to work to unlock the full potential of cloud-native analytics.

Introduction

An Evolving Cloud Analytics Landscape

The data landscape is quickly changing as companies migrate away from legacy solutions in favor of a more modern, cloud-based approach to data warehousing and analytics. A staggering 41% of companies are considering a move to cloud-based analytics in the next year as they look to cut costs, improve efficiency, and enable better decision-making at every level of business.*

The departure from legacy solutions shouldn't come as a surprise. Despite investing billions of dollars in the last decade on BI and analytics strategies, adoption has yet to clear more than 35% of all employees. Something is clearly missing.

> *"Pervasive business intelligence remains elusive, with BI and analytics adoption at about 30% of all employees."*
> *~Gartner*

Unfortunately, existing BI solutions have failed to make data accessible to business teams. In fact, 63% of employees report that they are unable to access data in the required timeframe to do their job.** This is largely because legacy BI tools rely solely on data teams to be the gatekeepers of corporate data, leaving out business users in the process. Today's BI software is difficult for business experts to use, lacks accessible data modeling, and tends to create workflows that leave business decision makers and data analysts frustrated and caught in never ending cycles of data and dashboard request queues.

As companies strive to become more data-driven, BI adoption is paramount. Organizations must invest in solutions to make data accessible for anyone in an organization, not just the data team. The good news: Snowflake's unique approach to cloud data warehousing has created new opportunities for companies to open up the data conversation to all with a cloud-native approach from Sigma that encourages curiosity, collaboration, and informed decision-making.

In this whitepaper, we discuss the inherent problems with the old ways of managing corporate data, and explore how taking a cloud-native approach with Sigma and Snowflake solve some of the biggest data challenges facing companies today. Read on to learn:

- The most common analytics challenges facing companies today
- How Snowflake and Sigma together provide the leading cloud-native analytics solution
- How to connect Sigma to your Snowflake cloud data warehouse
- How to optimize your Sigma experience with Snowflake

*Source: BI and Analytics in the Age of AI and Big Data, TDWI 2018
**Source: Computing Curiosity Gap, ResearchScape 2018

Common Analytics Challenges

Concurrency, Performance, and Scalability

There is more data generated than ever before. As companies grapple with unprecedented volumes of data, maintaining a high degree of performance when running parallel analytical workloads continues to be a challenge. BI teams are forced to schedule queries to run against the warehouse during off-peak times to manage limited scalability and concurrency of conventional data warehouses. The result: insights are delayed, business experts lack access to the data, and companies are faced with compromise on a daily basis; everyone is frustrated. Any database administrator should be concerned about scalability.

Meanwhile, conventional database architecture (e.g. shared disk and shared nothing) has failed to deliver a scalable solution to keep up with the rising demand for insights. These database architectures struggle to handle concurrent queries, partition data for multiple uses, and manage complex joins. This often results in reporting queues that takes hours or days to turn around, when it should take minutes. These issues only become exacerbated as data volume increases with business growth.
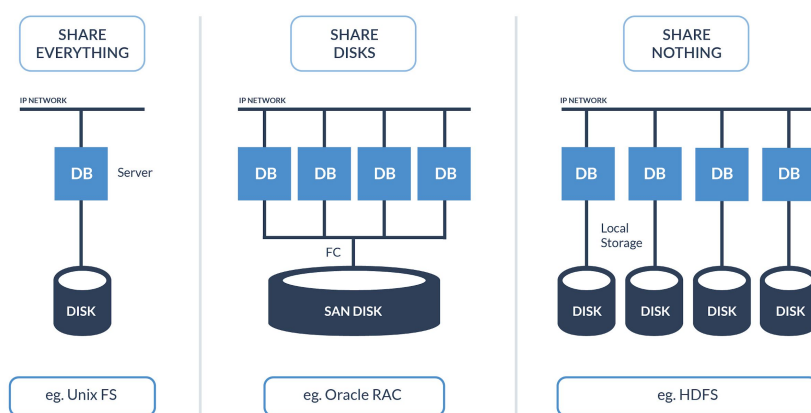

Figure 1. Common legacy database architectures

Flexibility

Scalability and performance issues of conventional data warehouses can't be ignored. So many companies try to offset the problem by provisioning warehouses to handle peak demand. This results in over-provisioned warehouses and constant management. Over-provisioning increases costs and is an extremely inefficient way of solving the problem. This inflexible approach can cost companies millions over time as database capacity goes unused.

Another common issue companies encounter with traditional data warehouses come from the limitations surrounding data support. Today's web, IoT, and API-driven world generates unprecedented amounts of semi-structured data (e.g. JSON, XML). Most legacy data warehouses cannot store semi-structured data, and those that do struggle to query it.

When it comes to analytics, legacy tools suffer from lack of flexibility as well. It can take months to set up data models, deploy reporting systems, or build dashboards using legacy tools. And when changes arise, or business needs evolve, the entire system can suffer from extended down time as data teams get taken off other projects to respond.
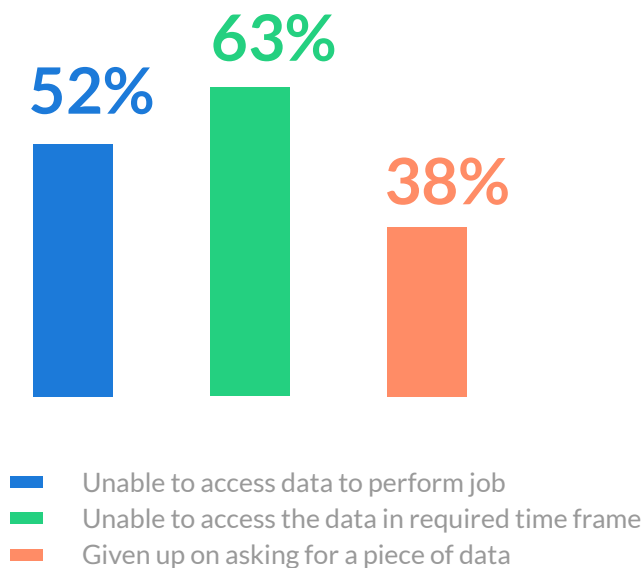
Data Modeling

Most analytics tools have built-in data modeling capabilities. But few products provide data admins with a flexible way to guide data exploration and build centralized data definitions without limiting business experts.

Traditional tools fail to deliver governance that balances the need for domain experts to find, add, and trust new data sources when conducting an analysis. Instead, they force data teams to invest weeks or months of time deploying rigid data models using SQL—some tools even require analysts and admins to undergo training to learn proprietary modeling languages. This approach to data modeling creates inefficient cycles of work for both the BI team and the business teams within an organization. Any time the data model needs to be modified to ensure the freshest and most relevant data is served up to business experts, someone has to write SQL—a skill set few business teams possess—to deploy changes.

These traditional modeling workflows are not only inefficient, but they have worked to keep business experts out of the iterative data discussion. As the business people closest to the data, they are the best suited to know what data is relevant to their questions. As BI teams work to update centralized data definitions or create base tables to kick off analysis and answer those questions, business experts wait days or weeks for model deployments. The result: 62% of business teams don't get answers in time to act in an informed way*.

Leading Obstacles to Data Insights

52% **63%** **38%**

- ■ Unable to access data to perform job
- ■ Unable to access the data in required time frame
- ■ Given up on asking for a piece of data

According to ResearchScape, access to data is holding back business users from generating useful insights.

Data Access and Collaboration

The lack of flexibility on the analytical side results in a gap between the business and data teams. Because these analytics tools require technical resources to manage, business experts are left in the dark throughout the entire process as they wait on data teams to curate data, refine data models, and deliver updated reports or dashboards that rarely tell the whole story.

Meanwhile, legacy analytics tools aren't intuitive or easy for domain experts to use. There are big learning curves that require SQL to ask any deeper follow up questions—leaving business users to rely on data analysts or stale data extracts that can be analyzed in Excel for answers. 50% of employees report using Excel instead of a BI tool*, because they lack SQL skills or real time data access.

Data Security and Governance

Data Breach Risk

Data breaches nearly doubled from 2016 to 2017, and 73% of all U.S. companies have now been breached*. The average data breach costs companies $3.86 million—up 6.4% from 2017**. The latest data suggest that 47%*** of these breaches were caused by human error or negligence, often stemming from lost or stolen data extracted to a spreadsheet. The Excel data extract workflow is ubiquitous in the enterprise at this point, but it presents significant and unnecessary risk to companies.

Not only are there high costs associated with managing a data breach, but companies must also consider how a breach can affect their customers. Typically, in the US data breaches tend to yield high customer turnover rates, which further exacerbate the impact on a business. This is especially true for digital businesses and can increase the average cost to $4.2 million on average**.

Data Visibility

Once a data set leaves the warehouse it's nearly impossible for a data team to track its whereabouts and ensure governance and compliance measures are met. Of course, business experts aren't looking to circumvent enterprise governance practices. They're just trying to get the answers they need to make better business decisions. And because they lack the extensive training or SQL programming expertise to work with data directly in business intelligence (BI) tools, they are often powerless to answer the questions raised in the last meeting or email. So they turn to what they know best: the spreadsheet.

On-Premises Security Challenges

Traditional on-premises data warehouses simply don't have the same level of security that a cloud vendor offers. Meanwhile, an on-premises warehouse requires constant governance oversight to ensure compliance with security standards. This requires dedicated personnel to monitor threats and keep data secure at scale. Keeping up with security standards and certifications is a full time job, and is particularly important for companies supporting customers in healthcare, finance, or the federal government.

Building and managing an in-house InfoSec team is challenging for any company, especially small to mid-size enterprises that lack the resources larger enterprises possess. And even with those resources at hand, dozens of large companies reported breaches in 2018—the largest being Marriott which reported 500 million users had data stolen from corporate servers last year. No company is exempt from the threat of a data breach.

*Data Breaches on the Rise, Electronic Privacy Information Center, 2018

**2018 Cost of a Data Breach Study, IBM, 2018

*** State of the industry Report, ShredIt, 2018

Sigma+Snowflake

A Match Made in the Clouds

Cloud Vision, Performance & Pricing

Cloud-First Approach

Legacy BI hasn't made good on the promise to make data easily accessible, or leverage the emerging power of the cloud. Sigma and Snowflake share a common cloud-first mission to change that. Together, they make the world's first cloud-native analytics stack.

Multi-Cluster Architecture Performance

Snowflake's multi-cluster, shared data architecture (see Figure 2 below) separates compute resource scaling from storage resources. This enables seamless, non-disruptive scaling. Snowflake processes queries and tasks in a fraction of the time conventional on-premises and cloud data warehouses require. Its columnar database engine uses advanced optimizations, including automatic clustering, which removes the headache of manually re-clustering data when loading new data into a table. Snowflake automatically scales to support any amount of data, workloads and concurrent users and applications without requiring data movement, data marts or data copies.
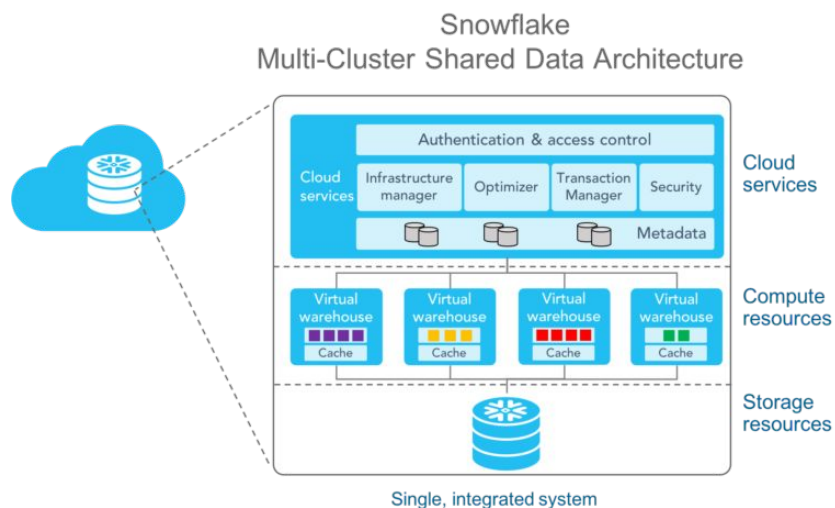
Figure 2. Snowflake architecture

Analytics Built for the Cloud

Sigma is designed to close the gap between Snowflake's data warehousing innovation and existing tools by leveraging the full power of Snowflake's cloud architecture. Sigma's cloud-native design allows users to take advantage of Snowflake's elasticity, compute, and scale. Sigma connects directly to the Snowflake data warehouse and can analyze billions of rows of data with ease. This cuts long query times, reduces reporting queues, and ensures that business users have a single place to explore mobile, web, and IoT data without having to become an expert in other SaaS tool —such as Salesforce, Marketo, and Workday—that serve as key data sources for the cloud data warehouse.

Usage-Based Pricing

Unlike other data warehouses, Snowflake has usage-based pricing for compute and storage, meaning customers only pay for the amount of data stored and the amount of compute processing used. This eliminates upfront costs, over-provisioned systems or idle clusters consuming money. And customers don't have to pay for each query or provision for peak loads like competing cloud data warehouses.

Sigma brings a similar cloud pricing model to the table. Instead of site or individual user licenses offered by traditional BI providers, Sigma provides unlimited access to data consumers, or "readers". An organization only pays for active users that perform interactive analyses against Snowflake. Accessing reports, dashboards, and worksheets are free for any employee, making it easy to get insights to those who need it without having to worry about costs. This innovative pricing model provides businesses with the ability to scale analytics across the organization and involve everyone in the data conversation at lower costs than traditional BI vendors.

> " *Sigma is like working in a spreadsheet with the power of Snowflake—giving our business experts access to the compute power of the cloud but with the freedom they'd get from a spreadsheet. Sigma opens up the true value of our data.* "
>
> *- Aash Anand, Data Analytics at Lime*

Analytics Without Limits

Sigma provides an analytical front end for Snowflake using a visual spreadsheet-like interface that is familiar—making data discovery easy. This allows business experts to pick up Sigma in minutes, with little to no training. The intuitive UI is powered by Snowflake and provides 99% of SQL functionality with zero coding. Anyone can perform analyses across massive data sets and build queries faster than any human can code using its SQL function library. All visual interactions and click commands inside Sigma are translated into SQL 'under the hood', unlocking true self-service analytics for business users that lack SQL skills. No query is off limits, no matter how complex.

Snowflake consolidates data on a single platform, whether extracted from traditional sources or newer machine-generated sources, without requiring cumbersome transformations and tradeoffs. Its patented technology natively loads and optimizes both structured and semi-structured data such as JSON, Avro, or XML and makes it available via SQL without sacrificing speed, performance, or flexibility (see Figure 3).

Sigma provides the ability to analyze Snowflake data in both structured and semi-structured formats. It's also possible to quickly flatten, parse and analyze modern semi-structured data (like JSON) in seconds.
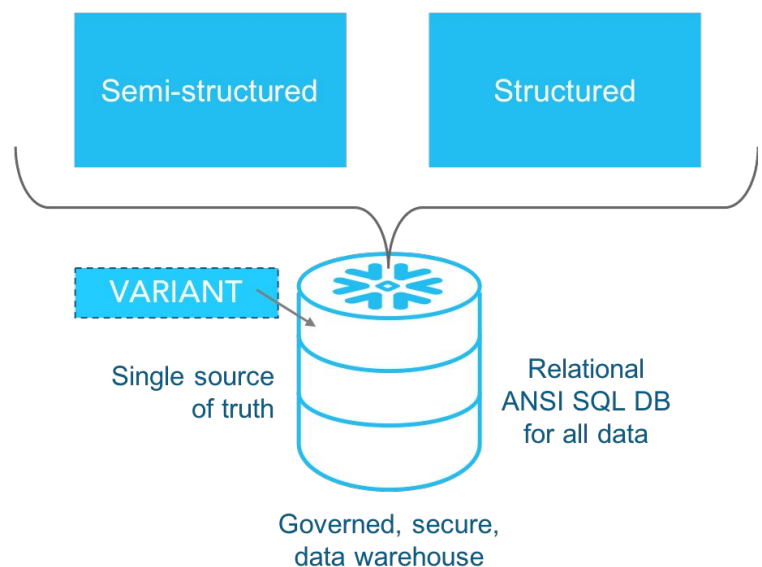

Figure 3: Consolidation of diverse data types.

A New Approach to Data Modeling

Visual Modeling Tools

Sigma's data modeling layer provides a flexible way to guide people's data exploration and build centralized data definitions without holding back business experts. It provides balanced control (as much as an organization requires) along with the freedom to let business users find, add, and trust new data approved by data teams. Sigma's central data definitions ensure consistent metrics across the organization.

Pre-Built Data Blocks

Data teams can curate starting points to accelerate analysis of Snowflake data with pre-built Data Blocks. Data Blocks can be constructed using the visual interface, without writing code. There is no need to learn new modeling languages or write SQL against Snowflake unless data teams want to use our SQL runner, which provides a warehouse index, auto-complete functionality, and SQL history—so analysts never have to search for past queries and avoid inefficient copy/paste workflows. Together, this provides a first-class experience for both the analyst and the business person in a single collaborative environment.

The visual aspect of Sigma's modeling capability opens up modeling to people outside the data and security team, meaning anyone Sigma admin can be a part of the iterative data discussion. And since business experts no longer have to wait for model changes they are less likely to download extracts and go. Your data remains in the warehouse where it's safe and secure.

Increased BI Adoption

Accessibility

Historically, analytics tools have failed to give domain experts the freedom to explore data. Instead, users were confined to reports or dashboards. Sigma has changed the paradigm. Now the data team can curate data for analysis or allow business experts to iteratively curate data themselves, while still maintaining a healthy balance of control where needed, unlocking the best of both worlds. This provides a single source of truth all teams can access.

Collaboration

Cloud apps like Google Drive and Slack have changed the way people work. Now business experts can get the collaborative features they love from their analytics tool too. Sigma supports any number of users, so everyone can access insights and explore Snowflake data. Even more, Sigma supports both individual workspaces and team workspaces, making it easy to collaborate on data analysis while still maintaining data permissions. This allows teams to collaborate, save time, and unlock the answers to any question, together in an iterative way. Team workspaces and Snowflake Data Sharehouses can also be shared externally, providing data access to partners or third-parties (in the form of a worksheet, dashboard, or report) that can benefit from data access.

Embedded Analytics - Reports and Dashboards

Sigma supports embedded analytics, making it possible to publicly or privately embed your Sigma visualizations, dashboards, and reports on a public web page, inside a company portal, or a third-party application. And because Sigma is connected directly to Snowflake, not a stale extract, the data is always fresh and up to date.

Cloud Security & Governance

Snowflake Data Warehouse Security

Snowflake follows best-in-class, standards-based practices to ensure data and data warehouse security. Security was built into Snowflake at the very beginning, not as an add-on or afterthought, removing the complexity and burden of enabling security from its customers. Snowflake provides role-based access permission, data encryption at rest and in transit, and utilizes private link connectivity so data is never transmitted on the open internet.

Snowflake customers can choose to run their cloud-built data warehouse on Amazon Web Services, Google Cloud Platform, or Microsoft Azure's highly secure infrastructure platforms. The Snowflake service runs inside a Virtual Private Cloud, with individual hosts protected by firewalls configured with the most stringent rules. All communication with the Snowflake service is protected at the network level using industrial-strength, secure protocols.

Snowflake leverages established best practices for security controls as part of its security program. And it works with AICPA-certified, third-party auditors to maintain security compliance and attestations including both SOC 1, Type II and SOC 2, Type II. Snowflake is also PCI DSS certified, FedRamp Ready, and HIPAA compliant, with the processes and controls in place required by the U.S. Health Insurance Portability and Accountability Act (HIPAA).

Sigma Data Security

Sigma + Snowflake is the BI solution built and optimized for the cloud. Data always stays in Snowflake—and off employee machines—where it's fresh, secure, and governed. This eliminates the need for desktop tools. Sigma sits on top of Snowflake, using a secure connection to query your data warehouse directly. Sigma gives read-only access, materialized views, and writes queries to answer questions that are returned in the browser. And data is never stored at rest.

Because Sigma provides a single point-of-access for your data, it's possible to establish robust data governance while still enabling business people to answer their own questions.

Sigma's approach keeps report sprawl to a minimum and access to sensitive information restricted. Administrators can set permissions by team and namespace, and can restrict data access directly from the database. Ultimately, this gives companies greater visibility and data access without the traditional risks encountered with other third-party analytics tools, such as Microsoft Excel or Google Sheets.

Sigma + Snowflake

Best Practices

Connecting Snowflake to Sigma

Connecting your Snowflake data warehouse to Sigma is simple. Unlike other BI tools, Sigma comes optimized to work with Snowflake 'out of the box'. This dramatically reduces the time it takes to get up and running with a proof of concept.

There are two ways to go about making the connection. You can either use Snowflake's Partner Connect Portal to connect in one click, or connect directly in Sigma. To get a complete walkthrough of the process, check out the short video tutorial below.

IP Whitelisting

If you have IP whitelisting enabled on Snowflake, be sure to whitelist Sigma's IP addresses. If you don't whitelist the IPs below, then data will not be visible through your Snowflake connection.

Sigma uses the following IP address to connect to your data: 104.197.169.18 and 104.197.193.23t

Connect to your Data: Snowflake

User Permissions

Sigma can see the databases, schemas, and tables granted to the user profile used to connect to Sigma. If you connect to Sigma utilizing Partner Connect, a Sigma user is created on your Snowflake instance. By default, this user has all of the public grants. If you would like to access additional data in Sigma, you'll need to add those permission grants.

Learning to Use Sigma

Sigma is easy to use and often comes naturally to those who are familiar with spreadsheet programs. But there are many unique features and capabilities Snowflake users will want to learn to extract the most value from Sigma. We encourage any new Sigma user, whether a data admin, analyst, or business person, to explore the Sigma Academy, which can be found on the Sigma Home Screen once you login. There is also helpful How-To Documentation available on our website, along with a Function Index and other learning materials such as videos, eBooks, and white papers.

Get started

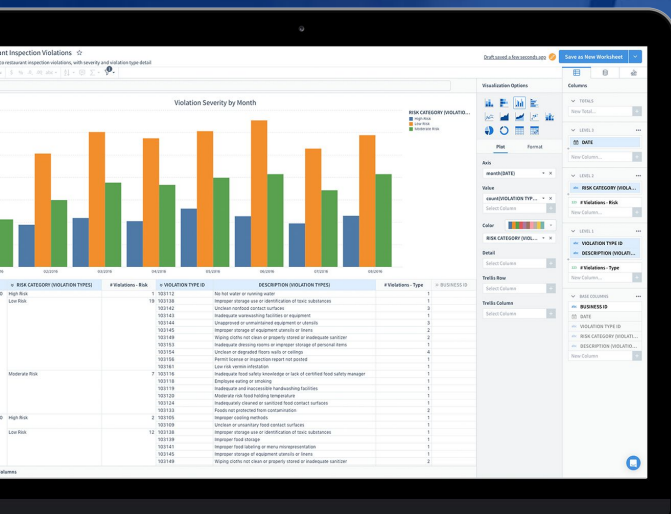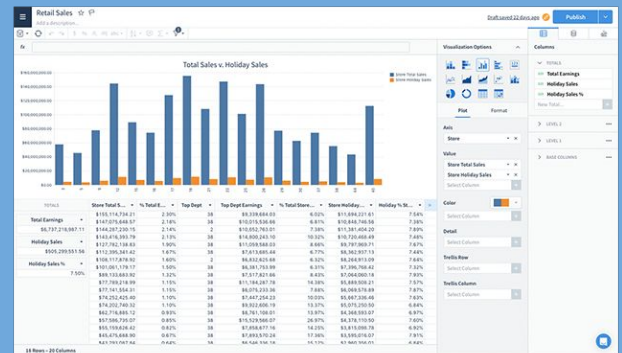Visit our [Website](#)