Sigma's Architecture

A scalable foundation for the future

AUTHOR

Phil Ballai
Enterprise Architect

INTRODUCTION

At Sigma, we're on a mission to empower everyone to make the best decisions—by eliminating the barriers blocking them from answers they need. We've reinvented and redefined what it means to work with data in the modern cloud era.

Customers have long been stuck looking at historical data trends through dashboards built by certified experts. Sigma is changing that. We're the only platform that scales to cloud volumes on live data, merging the simplicity of spreadsheets with the power of Python and SQL—and layering AI on top to turn everyone in the enterprise into a data analyst.

Our industry's first-and-only Input Table feature lets end users add their data to the warehouse, enabling anyone to build forecasts, perform reconciliations, and even create applications and workflows.

The world's largest enterprises and most innovative startups use Sigma to work with live data every hour of every day—not just finding insights, but forecasting the future and automating their workflows.

Sigma's Core Pillars

When Sigma was founded in 2014, we established several core pillars that define our approach to everything we do. Despite shifting technology trends, these pillars have consistently proven to be a winning model.

In this paper, we'll identify Sigma's core pillars and discuss how they shape our industry-leading product.

INSTANTLY FAMILIAR TO USERS	BUILT FOR THE MODERN CLOUD	UNPARALLELED SECURITY
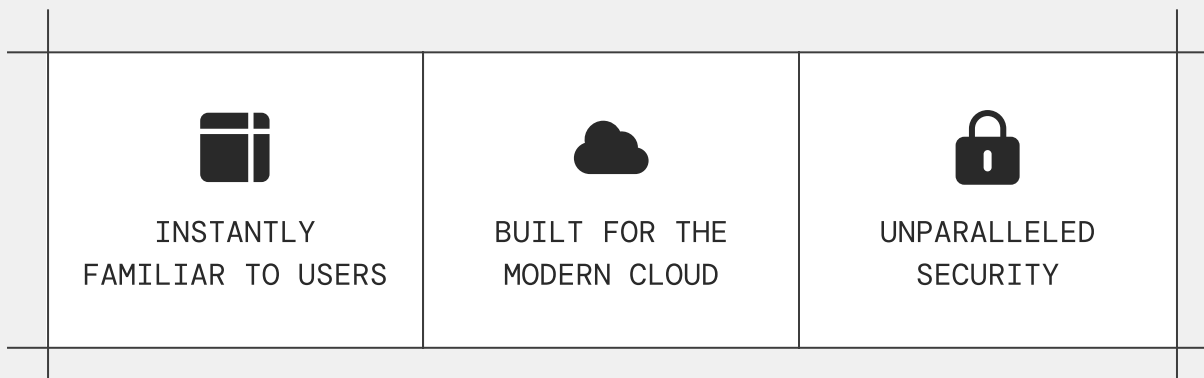
While it seems obvious, many products miss the mark in their design. There's a time and place for "breaking the mold" in product design, and sometimes it pays off big. Remember the first time you saw the Google search page?

But we're not introducing internet search to the world. We're catering to business users who've spent years working with spreadsheets or learning complex BI tools to get their jobs done.

With the advent of the cloud and massive data volumes, those old tools just can't keep up. But work still needs to get done. Users expect to interact with data in a familiar interface—a spreadsheet. They don't want to learn another tool or wait for a development cycle to get the information they need to make decisions.

Sigma is built for today's needs and users. With Sigma's intuitive spreadsheet-like interface, users can analyze, visualize, and collaborate—all at the scale and speed that business demands.

This maximizes productivity and ROI by enabling users to create value immediately, instead of having to be trained on some new tool(s) to work.
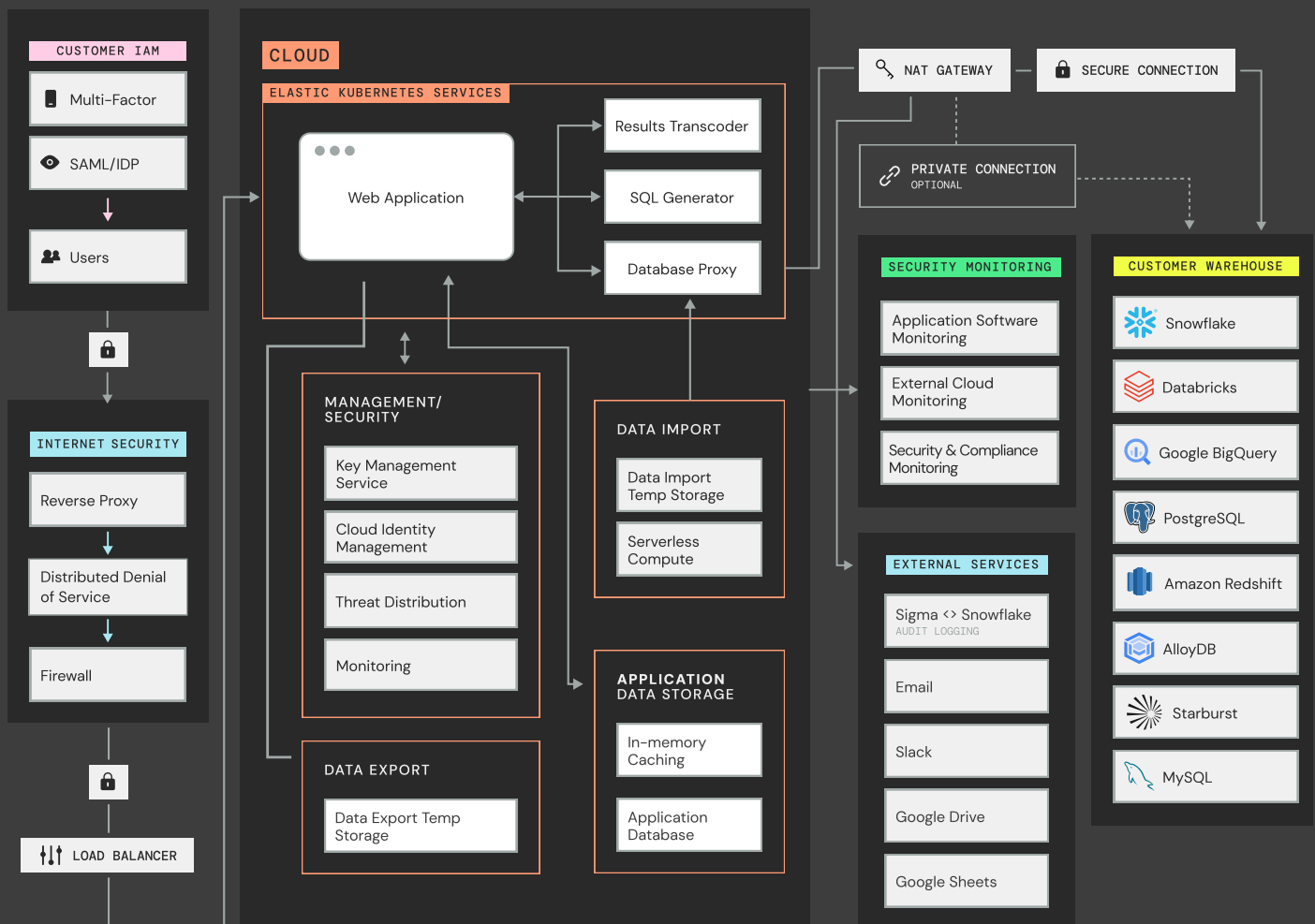
Sigma eliminates the need for on-premise software installation, seamlessly integrating with your cloud data warehouse. This allows us to avoid the burden of supporting legacy installations and dedicate our efforts to optimizing for the cloud.

Our architecture embodies a "best-of-breed" approach to SaaS design from end to end, as illustrated in the diagram below.

Some very important things result from this design approach.

sigma | Architecture

CUSTOMER IAM
- Multi-Factor
- SAML/IDP
- Users

CLOUD

ELASTIC KUBERNETES SERVICES

Web Application

- Results Transcoder
- SQL Generator
- Database Proxy

MANAGEMENT/SECURITY
- Key Management Service
- Cloud Identity Management
- Threat Distribution
- Monitoring

DATA EXPORT
- Data Export Temp Storage

DATA IMPORT
- Data Import Temp Storage
- Serverless Compute

APPLICATION DATA STORAGE
- In-memory Caching
- Application Database

INTERNET SECURITY
- Reverse Proxy
- Distributed Denial of Service
- Firewall

LOAD BALANCER

NAT GATEWAY

SECURE CONNECTION

PRIVATE CONNECTION
OPTIONAL

SECURITY MONITORING
- Application Software Monitoring
- External Cloud Monitoring
- Security & Compliance Monitoring

EXTERNAL SERVICES
- Sigma <> Snowflake AUDIT LOGGING
- Email
- Slack
- Google Drive
- Google Sheets

CUSTOMER WAREHOUSE
- Snowflake
- Databricks
- Google BigQuery
- PostgreSQL
- Amazon Redshift
- AlloyDB
- Starburst
- MySQL

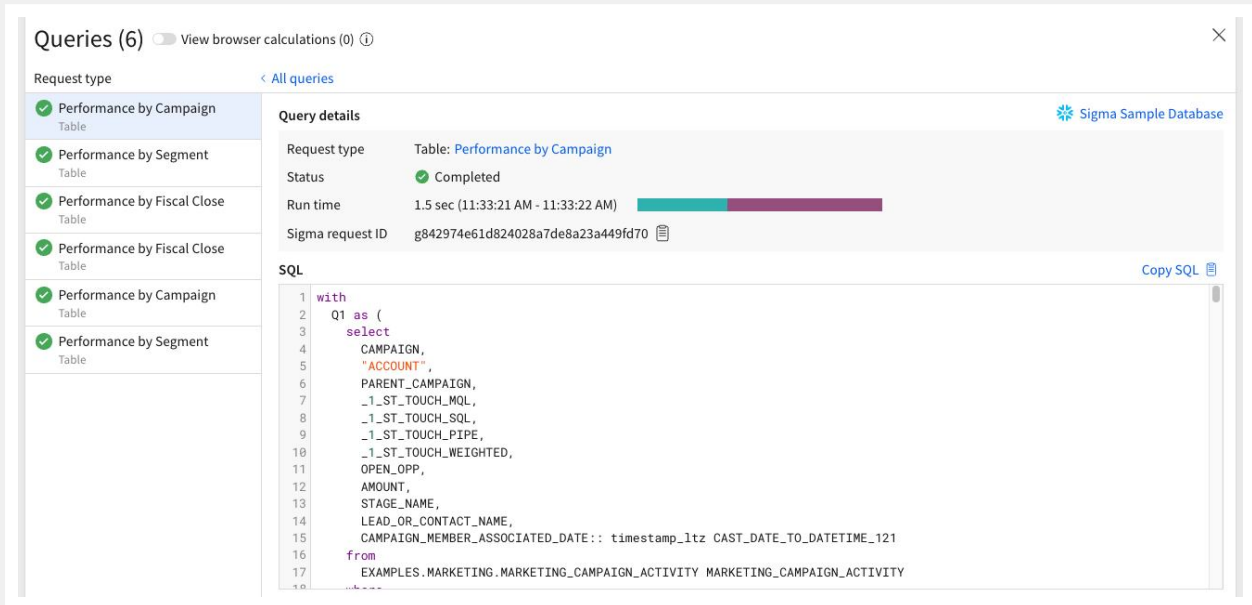## 1. Live Data:

When you access data in Sigma, it automatically generates optimized SQL tailored to your cloud data warehouse (CDW).

While most users won't need to think about the details, Sigma offers full transparency with complete visibility into every query it sends to your CDW.

Sigma sends queries over customer-administered connections, leveraging the robust security of your CDW with role-based access.

For enhanced security, Sigma also supports dedicated connections via Private Links.

The compute tier of your CDW ensures the data you receive is always current. After all, making business decisions with stale data isn't an option, is it?
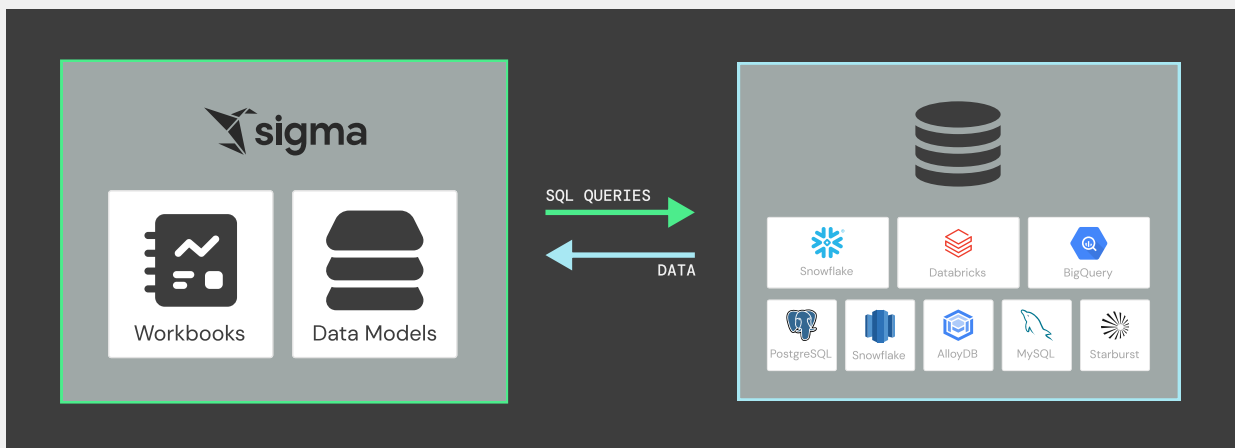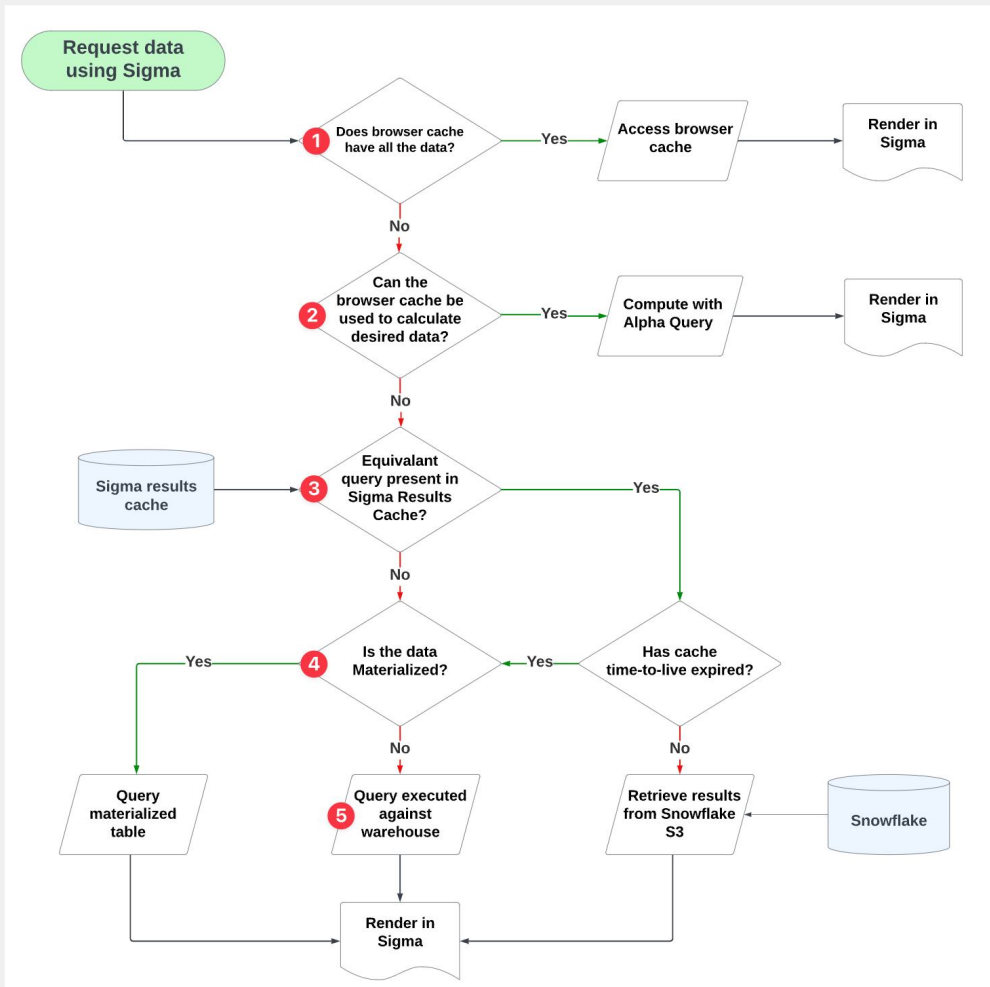
2. Massive Scale

Sigma effortlessly processes billions of rows in seconds, delivering unparalleled performance at scale for our customers. While many vendors simply scale up the warehouse compute instance to tackle processing demands, Sigma takes a more intelligent approach.

There are several ways this happens and implementations will vary based on the needs of the business.

Before we discuss point-specific features, it is critically important to understand that every Sigma customer will automatically benefit from the way Sigma operates when a user queries data.

How Sigma Performs Calculations

The diagram below provides a high-level overview of what happens when a user queries data in Sigma.

2. Massive Scale (continued)

1. When you query data in Sigma, we first check your local browser cache for the fastest possible response. This isn't an option the first time you query or if your cache has expired.

2. Once the data is cached, any manipulations you make—like adding a new column to calculate profit from price and cost—are handled using the cached data. No need to bother your cloud data warehouse (CDW) for those calculations. You'll see this in your workbook's query history as a browser calculation, which runs instantly.


3. Imagine two users viewing the same Sigma workbook simultaneously. The first user queries the CDW and retrieves the data. When the second user makes the same query, Sigma checks the warehouse results cache. Since the data was recently accessed, the query results are already cached in the CDW, not in Sigma. Typically, the CDW stores a copy of the results for 24 hours. The cache duration setting determines how long these results are kept, specified in minutes.

4. Sometimes, customers choose to precalculate complex data and store it using materialization. When this data is available, Sigma leverages it for your requests, boosting the speed and performance of your reports. Materializations improve the speed and performance of reports.

5. If the first four steps don't apply, Sigma queries the CDW to fetch the data you need.

2. Massive Scale (continued)

Sigma Alpha Query maximizes your CDW investment by efficiently handling massive datasets. Not only do users experience a significant boost in performance, but Sigma Alpha Query also continuously reduces the load on your CDW. Plus, as you add more users to Sigma, you'll typically see your cost-per-user go down.
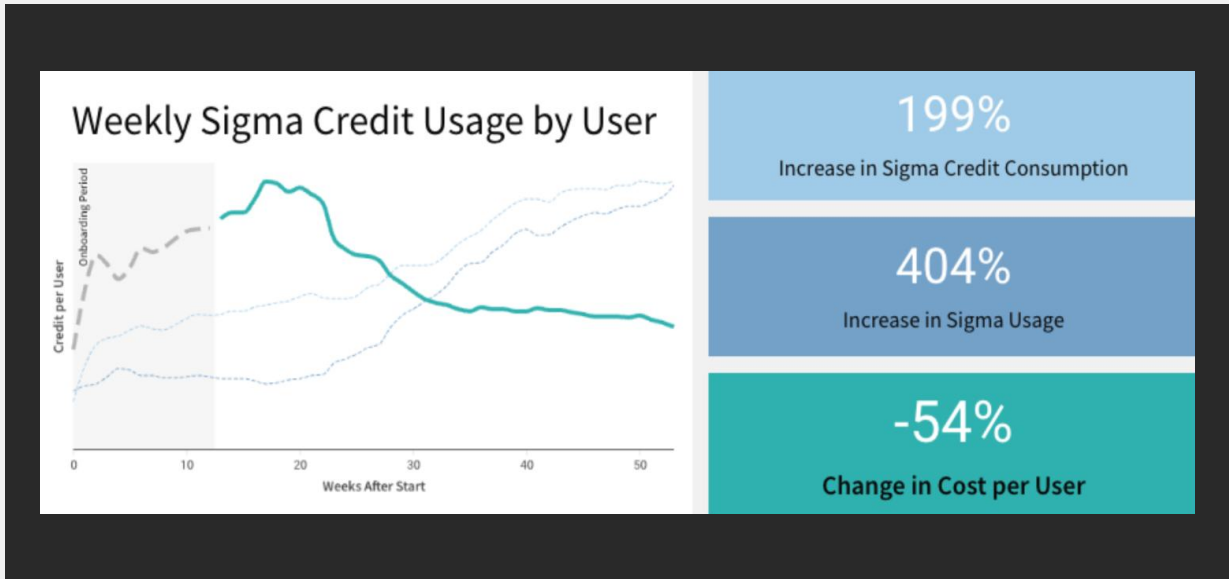
3. Multiple Performance Levers

Sigma offers unmatched flexibility, empowering each customer to discover their ideal setup. To design a high-performing system, it's essential to consider key factors we call "performance levels." These are areas to explore when your system isn't meeting expectations. We categorize them into three distinct groups.

WAREHOUSE LEVERS:
- **Snowflake:** Cluster Keys, Search Optimization, Warehouse Size
- **Databricks:** Liquid Clustering, Hive-Style Partitioning, z-Order Indexing, Warehouse Size
- **BigQuery:** Partitioning, Clustering, Auto-Scaling Capacity Slots
- **Redshift:** Distribution & Sort Keys, Data Compression
- **PostgreSQL:** Indices, Partitioning, Configuration Parameters (shared_buffers, work_mem)
- **MySQL:** Indices, Partitioning, Storage Engine (InnoDB vs. MyISAM), Configuration Parameters (innodb_buffer_pool_size, query_cache_size, max_connections), Sharding

DATA MODELING LEVERS:
- Avoid Selecting All Columns
- Use Tables Not Views
- Materialized Models (for complex joins)
- Incremental Materialization (large datasets)
- Denormalized Model

SIGMA LEVERS:
- Filter Clause in Inner Selects
- Default Limit Clause
- Materialized Datasets (for complex joins)
- Drive Filters from Dimensions (not Facts)
- Default Filters

These are just some of the key areas where we often find optimization opportunities. Sigma is here to help you unlock the full potential of your environment.

Naturally, this leads us to our next topic: monitoring.

4. Monitoring

Having the power to optimize performance and manage costs is invaluable, but it's just as crucial to measure these factors quantitatively. Sigma offers three built-in, no-cost methods for monitoring system performance.

First, leverage Sigma's built-in audit logging via the Sigma Audit Logs connection. These logs capture data related to user-initiated events within your Sigma organization. The data is stored in Sigma's Snowflake instance for a rolling period, with the option to export it to another system for extended retention.

Second, utilize Sigma Templates to evaluate warehouse usage. Sigma provides seven templates specifically for Snowflake, which can also be adapted for other cloud data warehouses (CDWs).

These templates deliver precise, in-depth, prebuilt analyses to help you understand your company's Snowflake consumption, activity, and performance.
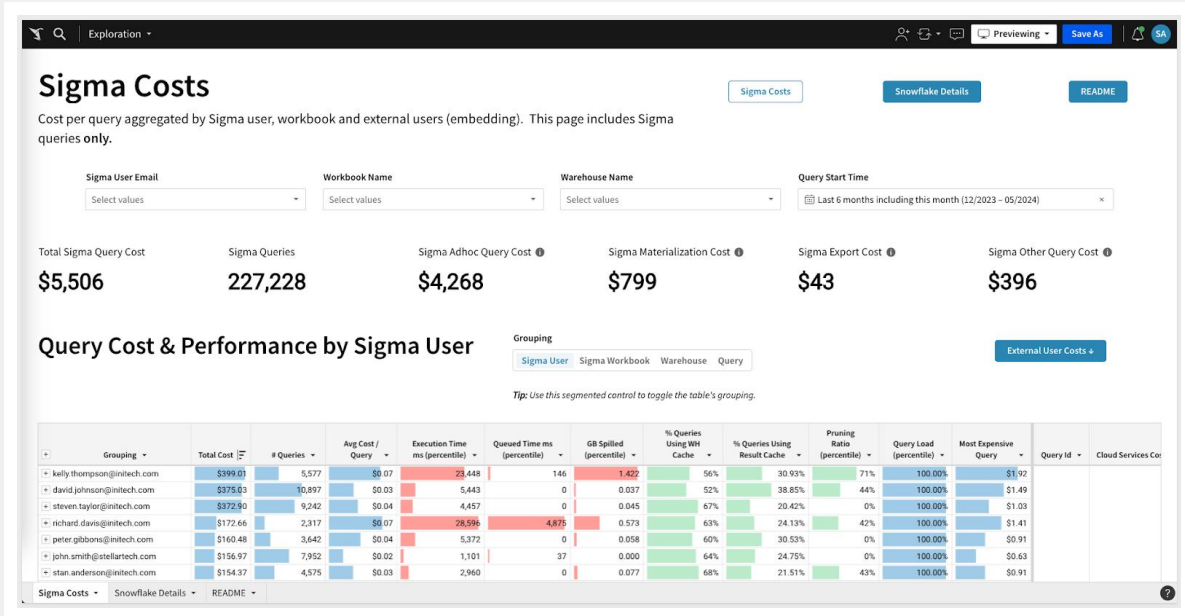
The seven templates are:

- Performance Analytics
- Cost Monitoring
- Cost Per Query
- Performance Monitoring
- Reader Cost
- Security Monitoring
- User Activity

4. Monitoring (continued)

Below is an example of the Cost Per Query template.

Third, access <u>Sigma's administrative "usage" dashboard</u> for a comprehensive view of your organization's Sigma usage and configuration.

One popular feature is the "Queries" report, detailing query performance and helping to identify users or workbooks with performance issues before they are reported.
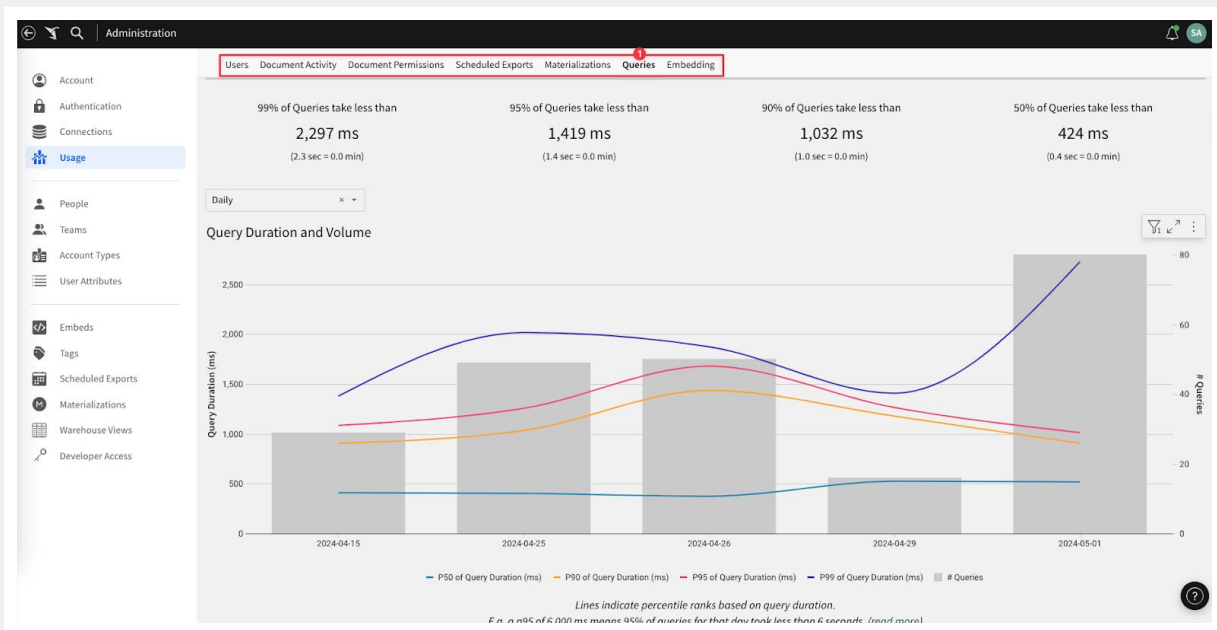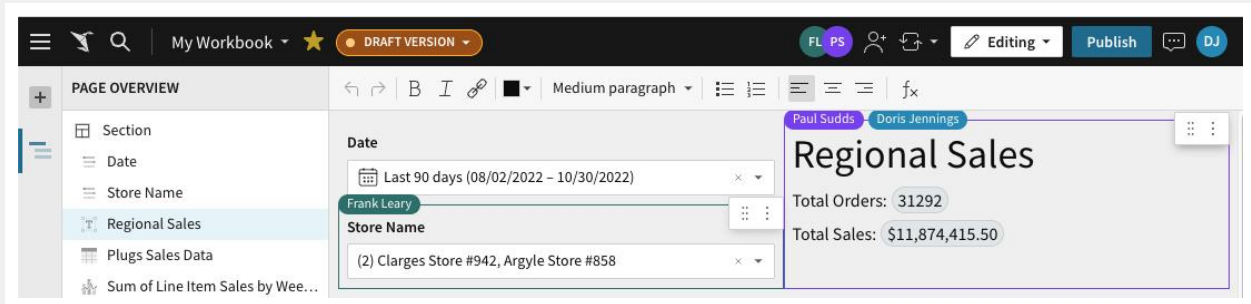
5. Collaboration

Sharing analytics the "old way" means emailing Excel spreadsheets or PDFs, or huddling in a conference room to review content. Sigma changes the game. While you can still share any content with anyone, Live Edit offers a superior approach.

With Live Edit, collaboration happens in real-time. Multiple users can work on workbook drafts simultaneously, with all editors sharing a single live draft. You can see and contribute to others' changes before they're published, no matter where you are in the world. Everyone is on the same page, working on the same thing, and seeing results instantly.

Sigma automatically shows the name of the user who selects an element in the top left corner of the selected element's bounding box.

This is truly a collaboration, as it was intended to be, getting the job done faster.

6. Cloud reliability

The cloud has proven its reliability across businesses of all sizes, making on-premise data centers a thing of the past. While there have been occasional outages, they haven't slowed the rapid adoption of cloud services.

Sigma partners with all major cloud providers—AWS, Azure, and GCP—and the leading cloud data warehouses. We ensure our systems are operational, redundant, and scalable to meet the daily demands of our customers.

REDUNDANCY	SCALABILITY
• Full zonal redundancy for each region • Regional disaster recovery	• Periodic Load Testing • Load Monitoring • Addition of Kubernetes Pods as Needed • Vertical Scaling of MySQL Database

At Sigma, we consider a secure architecture a fundamental necessity. We've made significant investments in this area, harnessing best-in-class security practices to safeguard your information.

Our platform incorporates features like immutable hosts, rigorous container checks, and advanced threat detection. We're dedicated to delivering secure data analytics, ensuring that every layer of our platform remains fortified.

Our Trust Center offers customers access to Sigma's latest security reports and documents, including ISO certifications and our SOC 2 report.

Sigma has invested significant time and resources in achieving compliance with the following standards:

Compliance

CCPA · EU-US DPF · GDPR · HIPAA

ISO 27001 · ISO 27001 SoA · ISO 27017 · ISO 27018

ISO 27701 · SOC 1 · SOC 2 · SOC 3

For more information on Sigma's commitment to compliance, click here.

Alongside a secure architecture, it's crucial to emphasize that customer data never leaves the Controlled Data Warehouse (CDW), ensuring its ongoing security, governance, and freshness. Users operate within a familiar interface, eliminating the need for Excel exports.

This shift enhances data security and reduces errors typical in spreadsheet usage.

Sigma retains necessary information temporarily to support user functionality. Customer data is held briefly, then promptly flushed from disk or cache upon completion of delivery.
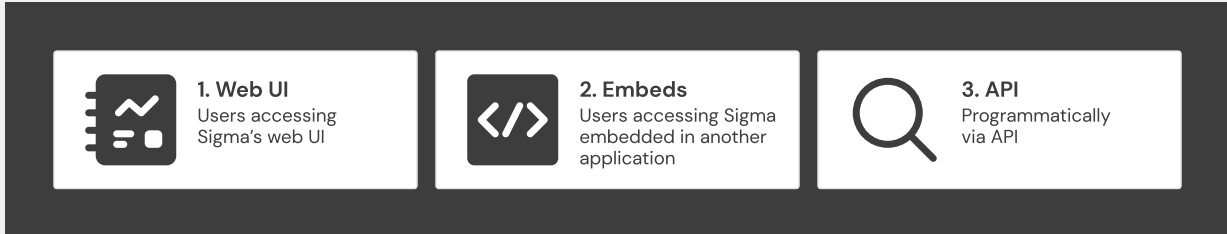
Extended storage pertains solely to workbook and system configuration details.

Items stored in the "Extended Storage" red box persist while the Sigma instance remains active and until users delete related items; all other data is short-lived.

SERVER MEMORY	EPHEMERAL DISK	CLOUD STORAGE	REDIS
• Decrypted query results (until delivery) • Input table entries	• Large query results • Excel exports via UI	• Exports via API • csv Imports via UI	• Query metadata for Query ID Cache

EXTENDED STORAGE	
• Warehouse connection details	• Workbook, dataset definitions
• Schema metadata	• Workbook, dataset version history
• Username, email, encrypted password	• User comments
• Team definitions	• Scheduled export definitions
• Workspace/folder definitions	• Audit logs

1. Authentication

Users accessing Sigma via a browser can do so using several methods:

1. Web UI	2. Embeds	3. API
Users accessing Sigma's web UI	Users accessing Sigma embedded in another application	Programmatically via API

1. Users accessing Sigma via a browser can do so using several methods:

Basic authentication: email and password:

- Sigma supports regular and guest users
- Multi-factor authentication is available

SSO/SAML: Single Sign-on using any compatible identity management provider:

- The SAML protocol enables access to Sigma using a single set of login credentials.
- It works by passing authentication information between the identity provider (IDP) (e.g. OneLogin, Auth0, OKTA, Azure, Google SAML, Ping, etc.) and Sigma
- Sigma supports any SAML provider that uses SAML 2.0

SSO/OAuth:

- The "Open Authorization Protocol" provides secure delegated access to applications with tokens, not passwords.
- Sigma also supports the key pair authentication method (public key + private key) for Snowflake connections.

1. Authentication (continued)

In the diagram below, we see how Sigma interacts with your IDP and Snowflake to authorize users' access to data.

1

End User tried to access Sigma

2

Sigma sends auth code along with client details to IDP for verifaction

3

IDP sends access / refresh token back to Sigma

4

Sigma Passes a connection string to Snowflake with OAuth token

5

Snowflake validates OAuth token

6

Snowflake performs user lookup

7

Sigma validates the signature in the URL using the same key as in step 3

Sign In To Sigma

email

password

SUBMIT

Identity Management

sigma

snowflake®

SAML authenticates the user's identity to a service (IDP)

Auth authorizes the user to access specific resources owned by the service provider.

Both can be used for single sign-on (SSO), which permits users to access IT resources with only one set of login credentials (e.g., username and password)

1. Authentication (continued)

2. In the second use case, a customer has embedded Sigma into their own application.

The process flow is as follows:

1. The customer's application authenticates users against the customer's authentication provider.

2. The customer's front-end application makes a call to an embed URL generation API hosted by the customer.

3. The Embed API generates a signed URL.

   a. The HMAC–SHA256 signature is created using a key obtained from a customer's Sigma instance.

   b. Each signed URL is single-use only. For details on how to create an Embed API, see here.

4. The signed URL is passed back to the front-end.

5. The front-end sets the URL as the "src" property of the iframe.

6. Sigma validates the signature in the URL using the same key as in step 3.

7. On verification, the embedded dashboard is rendered in the customer's application:
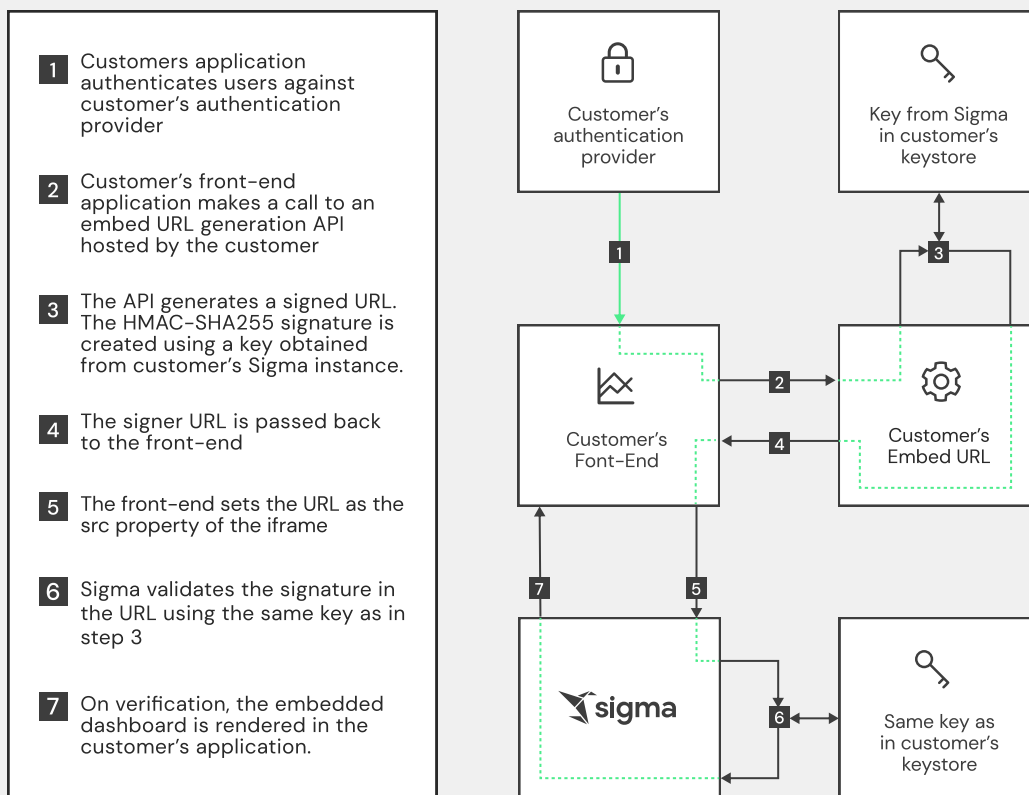
3. Sigma also provides a REST API by which customers can interact with Sigma programmatically. More information on using Sigma's API Reference can be found here.

2. Authorization

Authorization in this context is giving someone (or account) permission to have access to something.

In Sigma, this refers to (at a high level):

Content
- Workbooks
  - Sigma workbooks are data exploration tools for both Business Intelligence (BI) developers and spreadsheet-savvy analysts.
  - Can be shared to individual users or teams.
  - Can be saved into a folder or workspace, adopting that entities permissions.
- Folders
  - A storage container for Sigma content.
- Teams
  - A group of users who require common permissions.
- Workspaces
  - These are used to categorize and share folders and documents. Then they can be shared with users and/or teams using permission grants.

Data
- The connection to the CDW and related database, schema, tables and so forth that is accessible
- Row-level security
- Column-level security

Feature(s)
- For example, not allowing export of data. Sigma's role-based-access system provides granular control over what a user is permitted to do. To learn more about this topic, see here.
- To see the matrix of feature-based permissions, see here.
- In Sigma, users are assigned an "Account Type", either Lite, Essential or Pro.
- A custom option is also available and permissions of any account type can be adjusted, within the parameters of the license type.

3. Data architecture

Today, businesses leverage cloud data warehouses in diverse ways to meet their unique needs. For some, a single cloud account, database, and schema suffice perfectly—a streamlined approach at one end of the spectrum. Others integrate Sigma into their private-label applications, presenting several common approaches:

Sigma seamlessly supports these variations. Let's pinpoint and explore them further.

1. CO-MINGLED   2. SEGREGATION BY SCHEMA   3. SEGREGATION BY DATABASE   4. SEGREGATION BY ACCOUNT*
*SUPPORTED FOR EMBEDDING

1. Multi-Tenant with Row-Level Security (Co-Mingled):

- All customers' data is stored in the same Snowflake account.
- Multiple customers' data is stored together in the same tables, schemas, or databases but is logically separated to ensure each customer can only access their own data.
- Efficient in terms of resource utilization and cost.
- Requires careful design of security policies to ensure proper isolation.
- Sigma can address this in a few ways, depending on the business needs:

Row-level security (RLS):

- A calculated column is created in the dataset to compare the user attribute's value with the segregating column's value and return a true/false.
- A filter on the calculated column is set to return rows where the calculation returns true.

Calculated column(s) in a Dataset:

- Snowflake Row-Based Access Policies
    - A user attribute is created in Sigma to represent the Snowflake role.
    - In the Sigma connection configuration, this user attribute is set as the Snowflake role, thus parameterizing the Snowflake role Sigma uses to run queries.
    - Dynamic Role in Connection

3. Data architecture (continued)

2. Multi-Tenant with Schema-Level Isolation (Segregation by Schema):
- Each customer gets a separate schema within the same Snowflake account.
- Offers a balance between isolation and resource efficiency.
- Easier to manage than row-level security as each customer's data is separated by schema.
- Provides better performance isolation compared to row-level security.
- Sigma addresses this with:
    - A custom SQL dataset where the schema name is a parameter set as the user attribute allows Sigma to query the schema specified by the attribute's value.
    - Parameterized Queries

3. Multi-Tenant with Database-Level Isolation (Segregation by Database):
- Each customer gets a separate database within the same Snowflake account.
- Provides strong isolation similar to a single tenant but within the same account.
- Can be more manageable than maintaining separate accounts.
- Suitable for organizations where customers require significant data isolation but still benefit from shared resources.
- Sigma addresses this with:
    - custom SQL dataset where the database name is a parameter set as the user attribute allows Sigma to query the database specified by the attribute's value.
    - Parameterized Queries

4. Segregation by Account (Single Tenant):
- Each customer gets a separate Snowflake account.
- Provides strong isolation, ensuring that one customer's data and resources are completely isolated from others.
- Simplifies compliance and data security requirements.
- Can be more expensive and harder to manage at scale due to the overhead of maintaining multiple accounts.
- This use case is generally found in embedding scenarios.
- Sigma supports this method via:
    - Sigma can be told to query a different connection by setting the URL parameter to the desired value.
   - Dynamic Connection Switching

3. Data architecture (continued)

When selecting a tenancy model, it's crucial to weigh factors such as the necessary level of data isolation, compliance and security requirements, cost considerations, and the complexity of management and maintenance.

Each model comes with its own set of advantages and trade-offs, and the decision often hinges on specific needs and preferences. Sigma offers the flexibility to accommodate all these use cases seamlessly.
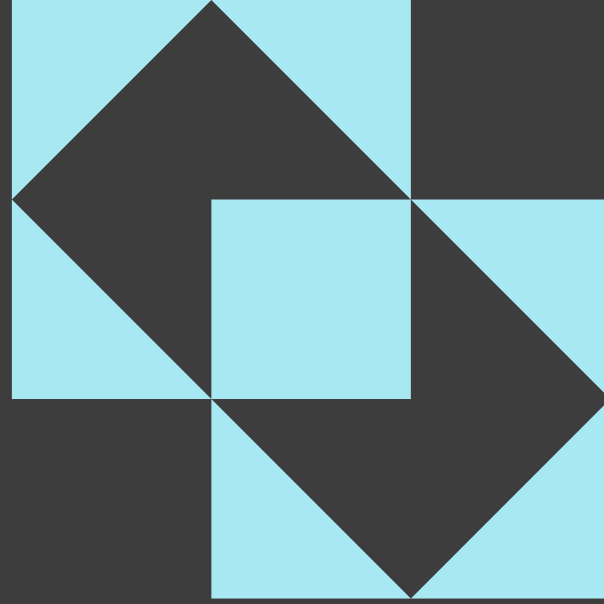
As shown in the image below, Sigma can accommodate the demanding use cases found in the modern enterprise today.
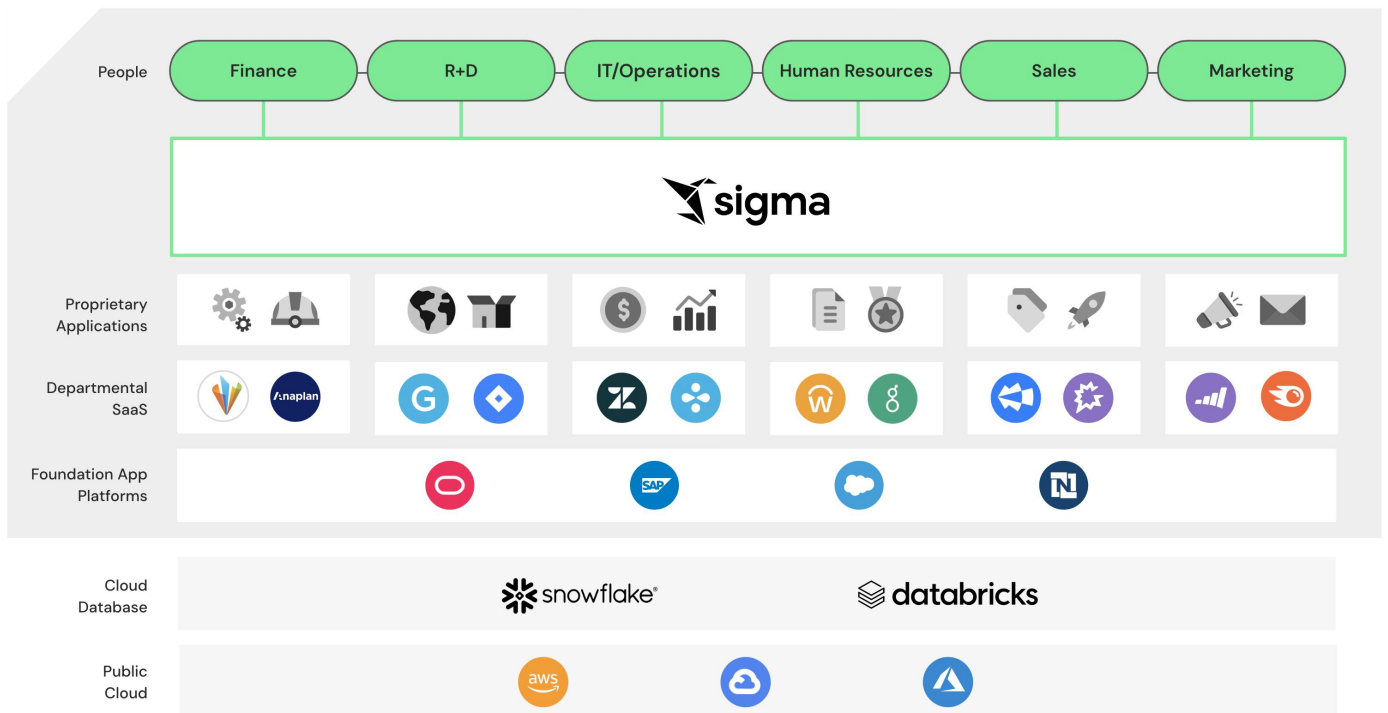
IN CLOSING

Sigma's is a modern cloud SaaS application that seamlessly integrates securely with your data warehouse directly. Users are able to access live data using an interface that is instantly familiar to them, so they can do real work immediately.

Simply put, Sigma is the UI for data!

Start a trial today

Help Center

Documentation

QuickStarts

Enterprise BI in a Spreadsheet UI

NEXT GEN ANALYTICS

Run granular business intelligence on billions of records without compromising speed or security–all from an intuitive interface anyone can use.