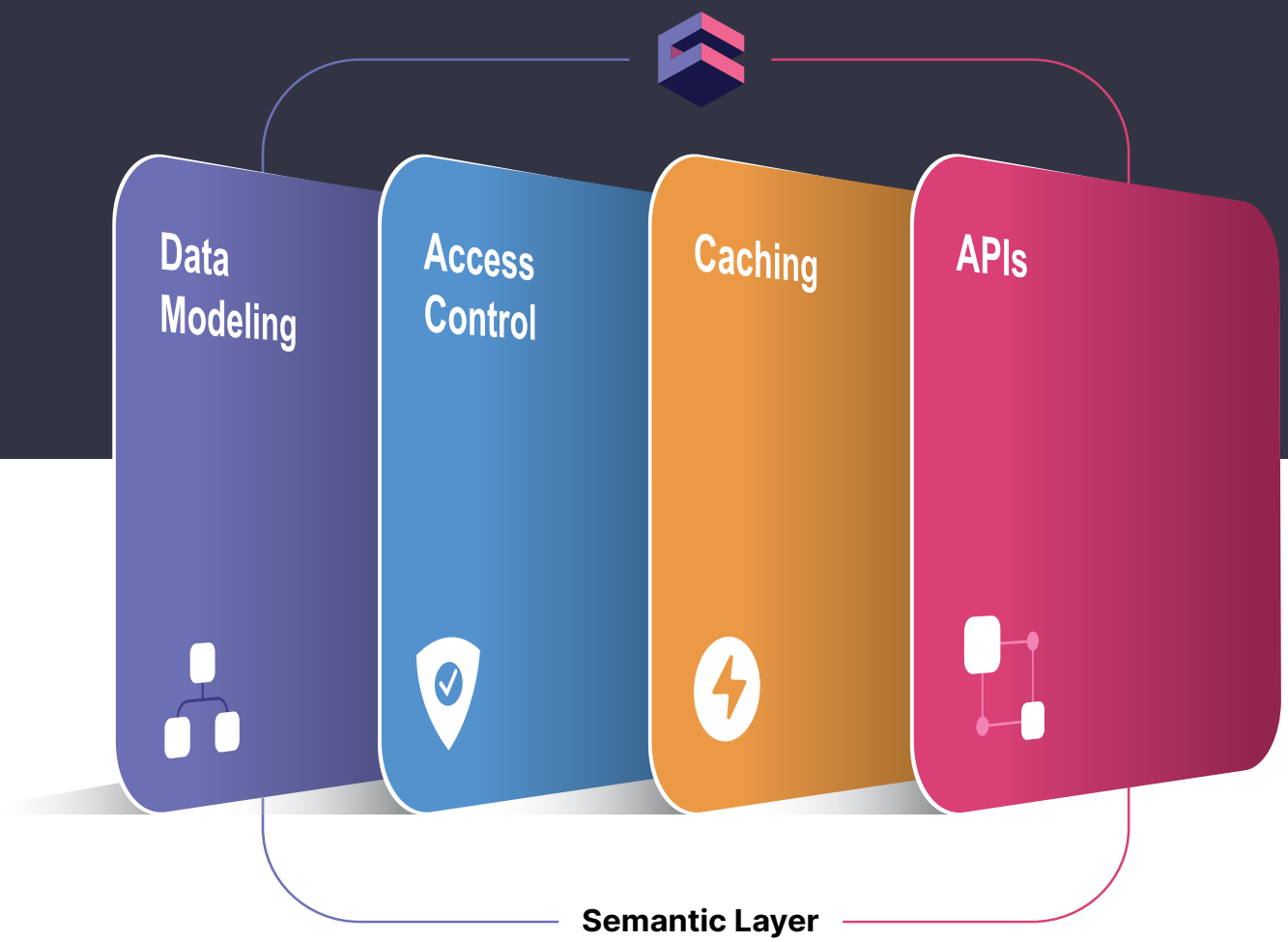
cube

The 5 Essential Features of Every Semantic Layer

Unlock the power of your Cloud Data Warehouse

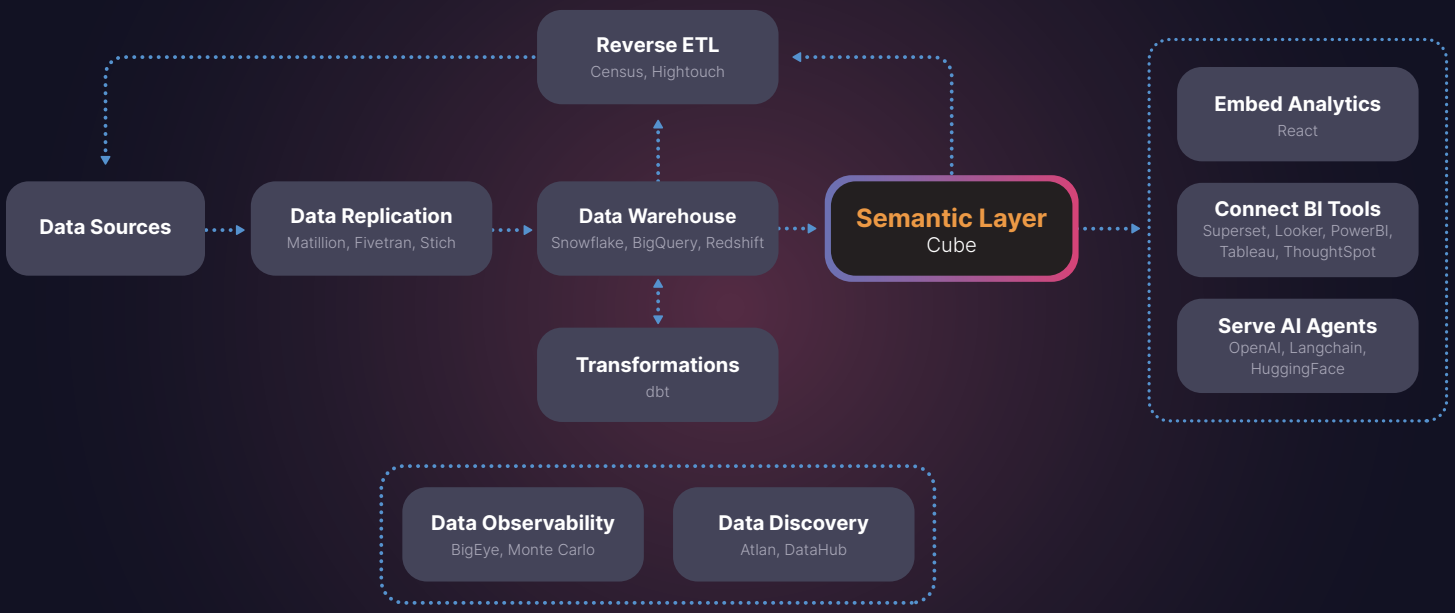


The Modern Data Stack Needs a Semantic Layer

Today data engineers have more comprehensive and powerful tools than ever before.

With efficient and scalable cloud data warehouses like Snowflake, AWS Redshift, Google BigQuery, or a data lakehouse like Databricks, the possibilities for fast, accurate, and meaningful data insights are immense.

However, with the power of these data warehouses comes the challenge of extracting the greatest value possible from your data. That's where implementing a universal semantic layer can take your analytics game to the next level.



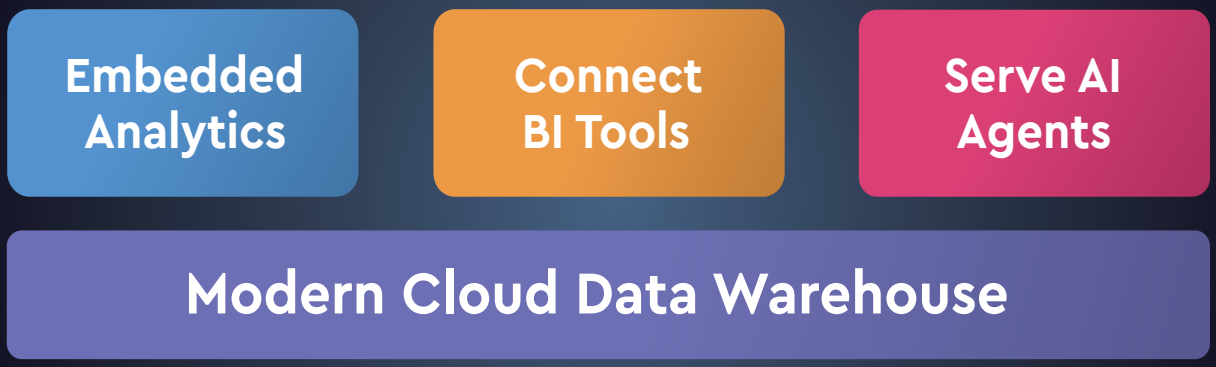
The universal semantic layer is more than just a buzzword; it's a way of efficiently organizing, structuring, and governing data to ensure that it is accessible, usable, and secure for any data experience that your business could desire. So what should you look for as you're evaluating various tools? There are 5 critical components to any semantic layer and as you review different tools, you should test them against these criteria.

1

Easily Deliver Data To ANY Data Experience

A universal semantic layer acts as a single source of truth for your organization's data, providing a consistent and clean dataset for users in various departments with different requirements. But once that data is clean and modeled, it needs to be delivered to a huge variety of data experiences.

Of course, any semantic layer solution should ensure seamless integration with popular analytics tools and data visualization platforms, including Tableau, Looker, and Power BI. But that is barely table stakes. In the data hungry world of today, you need to have the power to build brand new visualization experiences. Including serving AI agents and chatbots.



A perfect example would be embedding an analytics experience into your product to add functionality for your customers. Along with robust APIs, your semantic layer needs to have developer-friendly tools that make it easy to work in teams, building your data model collaboratively, with a sandbox for testing your ideas.

And while many out there try to build product-specific integrations, it is impossible to build all the possible connectors for the vast array of products that your users will demand.

A successful semantic layer must have mature and robust APIs that give you the flexibility to deliver anywhere – even AI agents.

RamSoft Delivers Beautiful In-Product Analytics

> We found Cube and, in two weeks, we delivered a highly customized and intuitive embedded analytics user experience – all without changing anything in our data stack.

Dhyan Shah
Product Manager at RamSoft

RamSoft

Dhyan Shah, Product Manager at RamSoft, a SaaS HealthTech company that provides medical imaging workflow, was looking for a solution that could serve as the basis of their in-product analytics to replace the traditional BI solution the RamSoft team had previously forcefit into their product. Like most BI tools, their BI software wasn't meant for embedding, with very little customization available beyond "changing some colors."

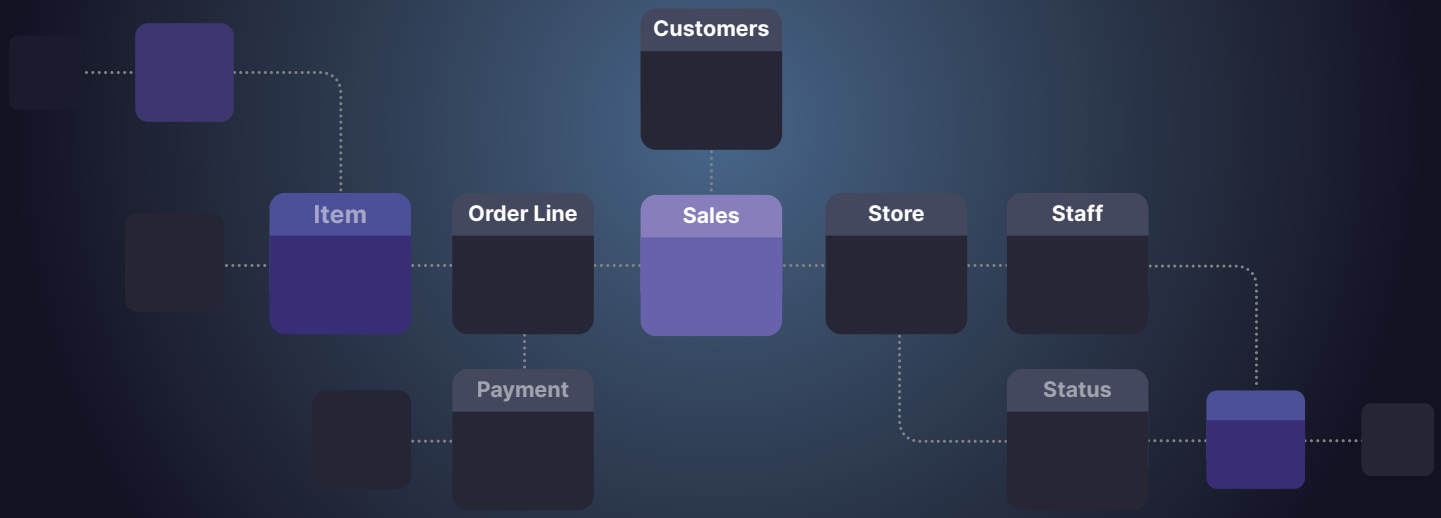
The team needed a platform to streamline the analytics behind the tool and allow them to focus on making the front-end seamlessly integrated with the UI of their product. They also wanted to customize the user workflows, report generation, and features available to users to avoid cluttering up the UI. After researching various options, Dhyan picked Cube, a semantic layer solution that could decouple modeling logic from the front end, allowing them to quickly build a completely custom UI in React. Using generated data schemas to get them started quickly, they took two weeks to build the data model, hook up their custom front-end with APIs, and deploy the semantic model into production – faster than anyone could have predicted.

The result? A highly-customized and intuitive embedded analytics experience that customers love.

2

A Single Source of Truth: Flexible, Centralized Data Models

The data model is the heart of the semantic layer and is usually the main reason companies start looking for a semantic layer solution. Data can quickly become messy and hard to manage without proper structure, organization, and governance. A universal semantic platform creates an abstraction layer over all your data sources, and should make it easy to create and manage your data models in a centralized location. This helps maintain the integrity and consistency of your data as your organization and its needs grow and evolve.



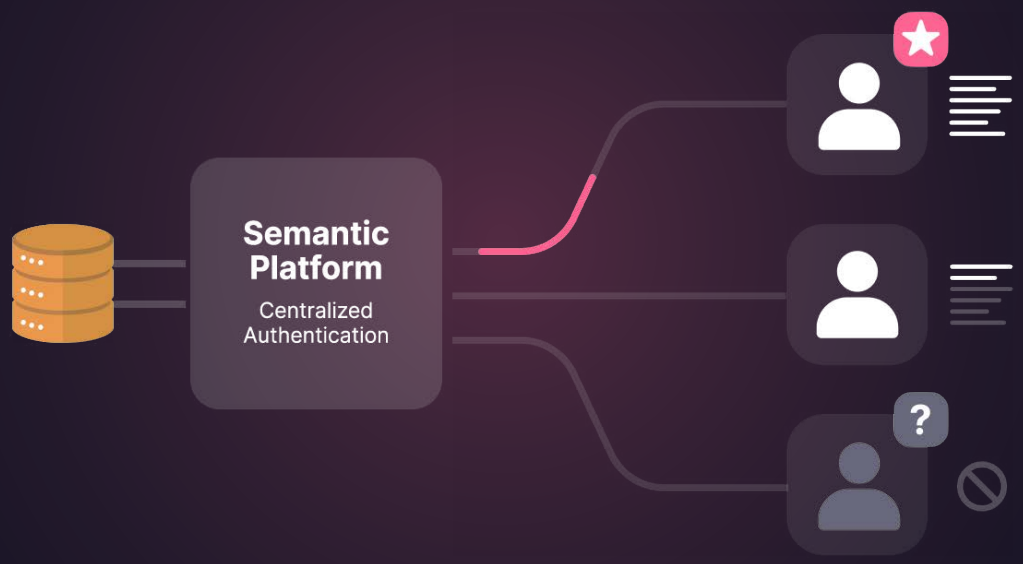
But in today's modern world, your data model needs to deliver a single source of truth to many different data experiences and centralizing your data models upstream makes sure that you only have to define them once. Whether you're powering a multiple BI tools, an AI chatbot, or an embedded analytics application, make sure that they'll all be powered by consistent and accurately modeled data – and skip repeatedly defining metrics for each presentation layer. The solution you choose should also have the ability to mirror the thin data models housed in BI tools and other data experiences to make it faster to get started.

With a centralized data model, you can ensure the integrity and accuracy of your data as changes are made. This way, each end-user experience is powered by the same source of truth without sacrificing availability or accuracy.

3

Centralize Authentication for Easier-to-Manage Security Controls

With great power comes great responsibility; data security and privacy are non-negotiable in today's increasingly digitized and data-driven world. A universal semantic layer provides the opportunity for centralized authentication and security controls, ensuring that you have full visibility and control over who has access to your data and what they can do with it. But not all semantic layer solutions have added this functionality, leaving you to manage authentication tool-by-tool. Not the most scalable solution.



By having upstream data access control, it makes it possible to grant consistent permissions to individual data consumers at both the raw data and the metrics levels. If you're looking to embed analytics within your software product, security is even more paramount and you should make sure that your semantic layer is built on the industry standard of IAAA: Identification, Authentication, Authorization, and Accountability. JSON Web Tokens, role-based access control, Auth0 support, and audit logs form the foundation of a secure stack with which to build modern data apps. You should also require SOC2 compliance as well as vendors willing to offer HIPAA and GDPR abilities for all PII and other sensitive data. All of these security and compliance capabilities need to exist in the semantic layer that is the core of your data stack

By implementing a single point of control for security and access management, you will minimize the risk of unauthorized access and data breaches.

4

Deliver at the SPEED of Data

In a world where data is being generated and consumed at breakneck speeds, timely and accurate analytics are integral to your business's success. A universal semantic layer has the opportunity to either slow-down the delivery of data to your analytical tools – or massively speed it up. Make sure your semantic layer tool employs pre-aggregations with a caching layer to accelerate the performance of your analytics, reducing query times, and delivering an amazing experience to your users.

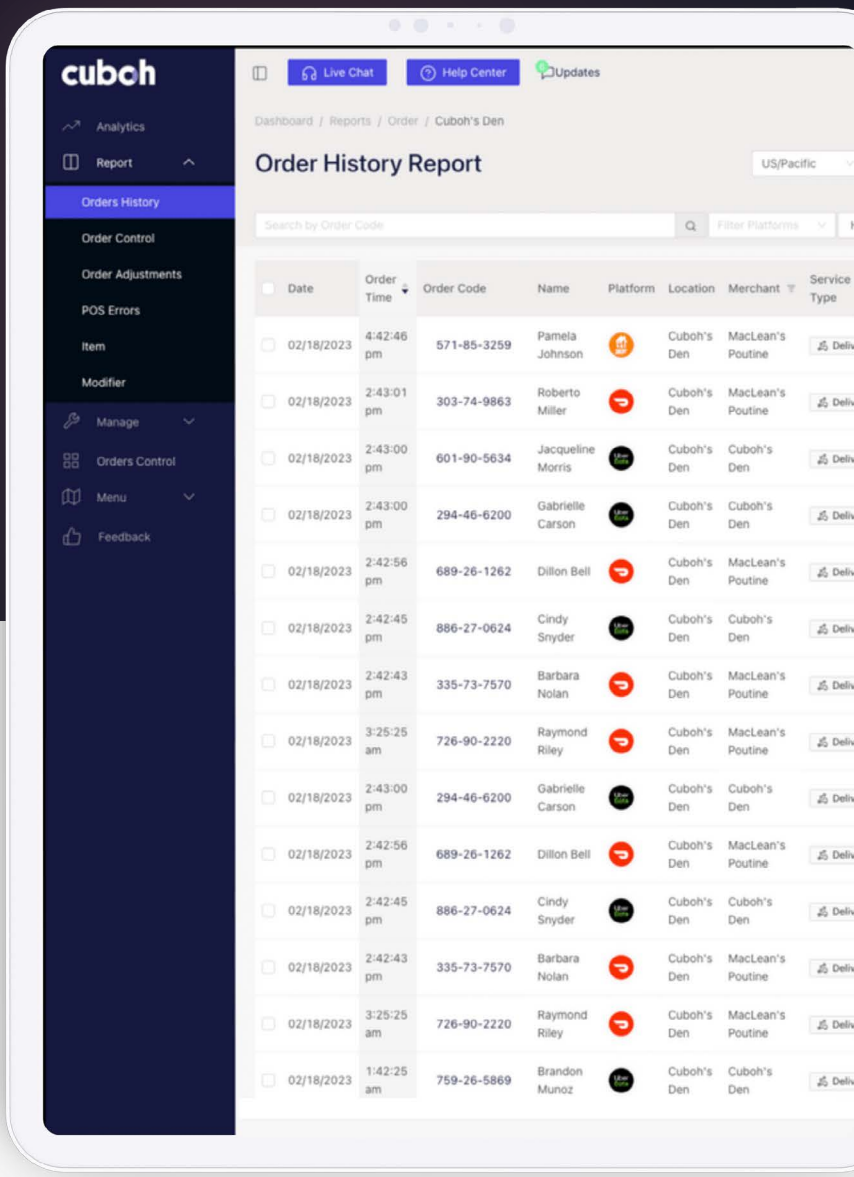
This caching system should feature in-memory cache and queue management capabilities that serve to buffer your database when there's an influx of concurrent requests. It should also support pre-aggregations, a layer of aggregated data managed by the semantic layer and kept fresh with a configurable refresh key. This pre-aggregation system also needs to be aggregate aware, meaning that your users don't need to change their query at all – the system will select the appropriate aggregation for them automatically. The features help maximize speed and minimize latency. A modern semantic layer can dramatically speed up the performance of the data experiences you provide to your users. Providing your customers or your business with faster and more reliable analytics will help you stay efficient and well ahead of the competition.

Cuboh Dramatically Speeds Up Analytics with Cube

> With Cube, we reduced the time required to generate real-time and historical reports from 10's of seconds to less than 2.

Brandon Ellis
Cuboh CTO

cuboh



Cuboh is a restaurant-tech company that makes it easy to integrate delivery apps with point-of-sales systems. They process millions of orders daily, but their data stack wasn't holding up. Querying the database took forever, and customers were noticing because they had to wait a long time for reports to process. They needed a caching layer that would integrate with their complex relational database and could pre-aggregate data to speed up query times.

After exploring different options, they decided on Cube because of its compatibility with relational databases, its robust Data APIs, and its advanced, two-level caching system.By implementing Cube, they were able to provide faster reporting to their customers in large part because of the pre-aggregation structure that made sure that fewer query metrics directly hit the database.

The result? Customer report requests went from 10s of seconds to less than 2 seconds.

5

Extensive APIs to Future Proof Your Stack

Once you've adopted a semantic layer, connected your data sources and modeled your data you'll want to begin using your metrics in all of your downstream data applications. Look for a semantic layer that supports your current, and future use cases with first-class application programming interfaces (APIs). Previous semantic layer technologies left your data stuck in your data warehouse which typically only speaks SQL over interfaces like JDBC and ODBC. These interfaces are not preferred for application development with the front-end frameworks your software engineers know and love. Instead, give them access to the same data consumed by business intelligence tools via preferred REST and GraphQL APIs. Modern semantic layers speak these languages as well as traditional SQL which keeps every application on the same page and reduces data management burden and distrust of metrics while accelerating time to value.

Evaluating semantic layer solutions

Make a list of companies that deliver semantic layer solutions so you can test out and trial their products for your specific use cases. Read reviews, understand their offering, and confirm that everything they say on the website is actually true.

It's recommended that you do a proof of concept with each semantic layer you are considering. Use this list of questions as a guide in your evalution

Can the product deliver data to any data experience such as embedded analytics, modern BI tools, and AI agents?	☐ Y	☐ N
How easy is it to model the data? Does the tool have a developer tool kit to streamline data modeling? Can it mirror thin data models that exist in many data tools?	☐ Y	☐ N
What kind of security controls does the semantic layer offer? Centralized authentification? What compliance certifications?	☐ Y	☐ N
Does the semantic layer have a caching layer that allows for pre-aggregations? Can it speed up queries and deliver an excellent experience to all data apps?	☐ Y	☐ N
Is there a community of data engineers to learn from? Does the company offer some or all of their offering in open source?	☐ Y	☐ N

Put Cube On Your List

A semantic layer with all the functionality you need

One company that you should include on your list is the open source company called Cube (cube.dev). Cube makes it easy to connect all your cloud data warehouses to as many data experiences as your employees, customers or partners could want – even powering AI agents and chatbots. Offering a universal semantic layer that not only centralizes your data model across many data sources, but also offers centrally managed authentication and security tools, a caching layer that can exponentially increase the speed of your analytics and robust APIs ready to deliver data to any data experience your users could want. Cube is the largest open-source semantic layer company with millions of downloads and a cloud offering that delivers additional tools, reliability, and high-availability. And you can try it for free, just connect your data sources to their free development instance to test out how it can work for your business.

- ☑ **Deliver data to any data experience with robust APIs**
- ☑ **Flexible data modeling with developer tools and easy collaboration workspaces**
- ☑ **Centralized authentication for access control**
- ☑ **Caching layer to optimize and speed up queries**
- ☑ **SOC2, HIPAA, GDPR compliant**
- ☑ **Supportive and active community of data engineers**

The Cube Mission

At Cube, we're on a mission to provide developers with analytical data access layer to build modern applications. Cube is used by thousands of companies around the world to power internal business intelligence tools, customer-facing analytics and give context to LLMs and AI agents.

Request a Demo
cube.dev