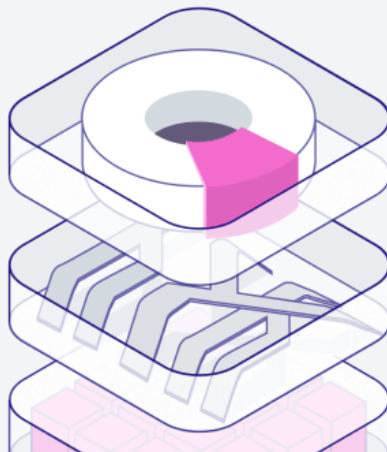




GoodData Architecture Overview



WHITEPAPERS

GoodData Architecture Overview

GoodData is a multi-tenant analytics platform that enables you to build modern, scalable analytics for your end users. GoodData's architecture is specifically designed for organizations striving to:

- Provide a solution that caters to all user personas — from business consumers to analytics engineers and developer teams who aim to build repeatable analytics workflows while applying software engineering best practices.
- Deliver and scale up an analytics solution to large numbers of separated teams, departments, or client companies (via a multi-tenant platform).
- Unify data analytics and business metrics organization-wide (capitalizing on a robust semantic layer).
- Build and manage data products; either as part of a SaaS solution or as an internal analytics solution (utilizing embedded integration and data-product management capabilities).
- Provide developers with tooling and integration features, and incorporate AI into the analytics.

Read the full whitepaper to learn how we empower organizations to achieve these goals with our platform's architecture and technical capabilities.

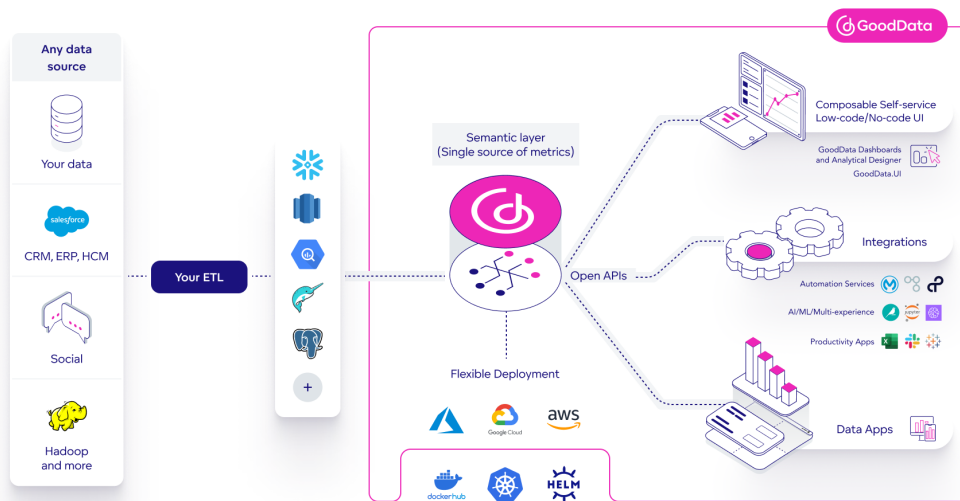
Whether a company has its own cloud data warehouse with all the necessary data for its analytical model or relies on multiple sources, the potential benefits of GoodData Cloud include:



CHALLENGES

- **Data access:** Connect your own data warehouses, or organize and manage data sources via data source managers to connect your data to the analytics.
- **Data refreshment:** Options to query data sources directly or minimize data operational costs through advanced caching technology.
- **Seamless integration:** Swift integration with preferred data tools via APIs and SDKs.
- **User-friendly self-service:** Low-code/no-code UI for AI-powered dashboards and charts, facilitating quick insights and decisions with features like Natural Language Query (NLQ).
- **Embed analytics anywhere:** Visualization and dashboard integration into apps, web pages, or data portals using iFrame, Web Components, or React SDK. Customize or white-label to match the company's brand.
- **Scale with multitenancy:** Analytics is provisioned to teams, partners, or customers in a way that ensures data separation and security within the existing stack.
- **Automation:** Analytics environment creation with collaborative code-based developer tools and blueprints for increased efficiency.
- **Data accuracy:** Governance features and a robust semantic layer for consistent, accurate, and secure data.

GoodData Cloud Architecture Overview



GoodData Analytics Overview

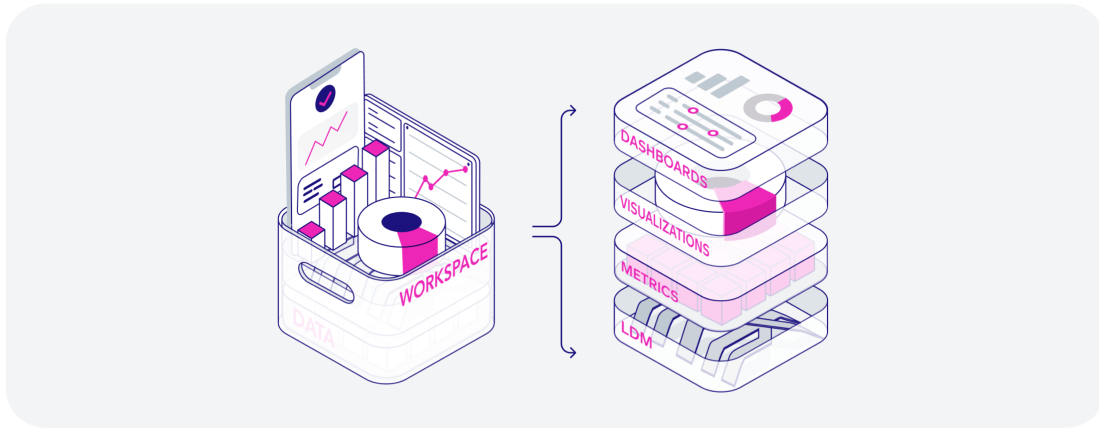
At a high level, GoodData's modular platform provides easy-to-use visual analytics tools, embeddable data visualization, and application integration. The platform offers robust multi-tenant features and thanks to its rich APIs and SDKs it is developer friendly and easy to customize. Our microservicebased architecture allows businesses to seamlessly integrate with existing infrastructure and effortlessly implement analytics solutions for employees, customers, or business partners at scale.

This whitepaper includes references to helpful [documentation](#) and [web pages](#). The aim is to provide an overview of GoodData's capabilities. If you would like to learn more, please [contact us](#).

Analytics Platform Features

Workspaces

A [workspace](#) is an environment where you organize, analyze, and present data for a specific customer, group of users or a specific use case. You can have one or multiple workspaces and control who has access to which workspace. Each workspace contains a Logical Data Model (LDM) and analytical objects including metrics, visualisations, and dashboards. Workspaces query your data warehouse in real time – no data needs to be loaded in. Caching can be used to optimize performance and lower data warehouse costs. Additionally, workspaces can contain localization and other customizations.



This workspace architecture is unique to GoodData and focuses on the following aspects required for managing large, complex analytical solutions:

- | Security and privacy:** Workspaces are completely independent and isolated from one another. If you are building customer-facing analytics for your product you would typically have one workspace for each of your customers. This ensures that each of your customer's data and insights stay private and a customer can only access data relevant to them.
- | Performance and scalability:** Smart caching reduces the query load on the source data warehouse, increasing performance and lowering costs. To scale your data solution, both provisioning and change management can be easily automated with GoodData. We will cover this in more detail later in the whitepaper.

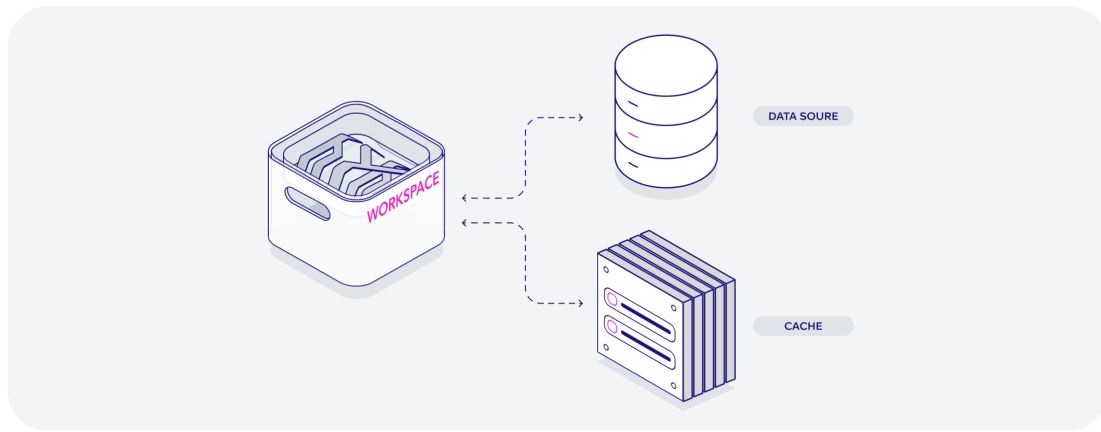
- | **Self-service:** GoodData workspaces offer powerful yet easy-to-use exploration and dashboard building tools. This allows users to gain deeper insights and make informed business decisions without needing help from data experts.
- | **Customization:** As mentioned, users are able to customize workspaces through building their own visualizations and dashboards. Beyond that, workspaces can be localized, styled, populated with custom visualizations, and more.

Data

GoodData lets you use your own [data warehouse or data source manager](#) and create a direct, real-time connection to your data. Once you connect a data source, GoodData scans your physical model to help you generate a suitable Logical Data Model and configure data mappings.

GoodData stores computed results for your analytics in an internal cache to avoid processing the same data by querying your data source over and over again. To notify GoodData that new data has been uploaded to the database, you can send a notification to the relevant data source. This trigger will invalidate the cache related to the data source so that the next time you open an insight, it is re-computed with the latest data from the database, and the cache is populated with the recent data.

With GoodData's caching technology, [FlexCache](#), built on [Apache Arrow](#), users can expect a significant reduction in data retrieval time. There are potential savings of over 50% on operational data warehouse costs, without compromising on the handling of increased data volumes or consistency of the user experience. You can calculate your actual cost savings [here](#).



Semantic Layer

Each workspace in GoodData contains a [semantic layer](#) which consists of an LDM and metrics. The GoodData semantic layer ensures that all of your end users, including self-service users, understand the data in the same way. The information in the semantic model can be leveraged for guided analytics and provides a shared understanding of the analyzed entities and their relationships. Objects that are created by analysts once can be used multiple times by other common users, helping them to interpret the data and perform ad-hoc data discovery. This concept allows users to retrieve valuable insights from their data, no matter their level of data literacy.

Logical Data Model

Each workspace has a Logical Data Model. This defines what data is available for analysis in that workspace and what the relationships of different data entities are. The best LDMs follow Star or Snowflake schema dimensional data modeling practices. Each object in the LDM has to be mapped to existing tables/views and columns in your data source. If a simple 1:1 mapping is not possible, GoodData allows you to write and save SQL queries to create derived tables and columns in the LDM without making changes to your data source. This feature is called SQL datasets. You can learn more about [SQL datasets here](#).

Benefits of the Logical Data Model

Setting up the LDM is a relatively small time investment, but it makes working with analytics and creating visualizations much easier down the line.

With a predefined logical model, you define the mapping and relations once, and then you (or even the end business users) can **reuse the same objects multiple times** for many different visualizations and dashboards. The system will make sure you don't combine things which should not be combined. This makes the LDM very useful for **self-service analytics**, and enables an easy to use yet powerful drag-and-drop visualization and dashboard building experience.

Having everything defined once and re-used, makes for **simple maintenance and change management**. For example, if you need to change a calculation that is used in 20 visualizations you only need to change it in one place, rather than changing it 20 times. Additionally, the LDM acts as an abstraction layer on top of your physical schema. This allows you to change the structure of the source data models without impacting the content you create in the workspaces.

The LDM is the foundation for everything you will build in GoodData above it: all the metrics, visualizations and dashboards. The LDM can always be changed and evolved at a later point.

Metrics

Metrics are reusable calculations built on top of the LDM. Metrics encapsulate business logic and are used in visualizations. The same metric can be used in many different contexts – sliced, diced, and filtered by different dimensions. This makes the system easy to use and provides consistency and data governance. Metrics are built using GoodData's proprietary Multi-dimensional Analytics Query Language (MAQL). GoodData computation engine then translates MAQL queries into optimized SQL.

MAQL (Multi-Dimensional Analytics Query Language)

MAQL provides dozens of analytical functions and operators, ranging from simple aggregation and filtering to advanced statistical and predictive functions.

Average Price of Products:

```
SELECT AVG (Price)
```

Revenue by a specific region:

```
SELECT SUM (Sales revenue)  
WHERE Region name = "West"
```

The syntax is reminiscent of SQL but there is no need for the 'join' and 'group by' clauses found in SQL. The analytics engine will automatically infer these from the LDM. Thanks to this, the metric definitions are simpler than if SQL was used. Metrics can also use other metrics, allowing analysts to break down the complexity of business logic to lower-level metrics with simple definitions.

Metrics are context-aware, which means you will not need to create duplicate metric expressions for specific combinations of dimensions in a visualization. You can create one metric and re-use this metric across different insights. Again, this is possible due to the business and analytical contexts defined within the LDM. The GoodData analytics platform automatically generates SQL queries that yield the correct calculation in a specific context.

Revenue:

```
SELECT SUM (Price * Quantity)
```

Cost of Sales:

```
SELECT (SUM(Cost of shipping) + SUM(Cost))
```

Absolute Profit Margin:

```
SELECT Revenue - Cost of Sales
```

Reusable Facts.

Declared in the LDM.

Reusable metric for many insights.

Clear semantics for business users.

New reusable metric for any insight.

Easily, logically built from core metrics. Context-aware when used with dimensions.

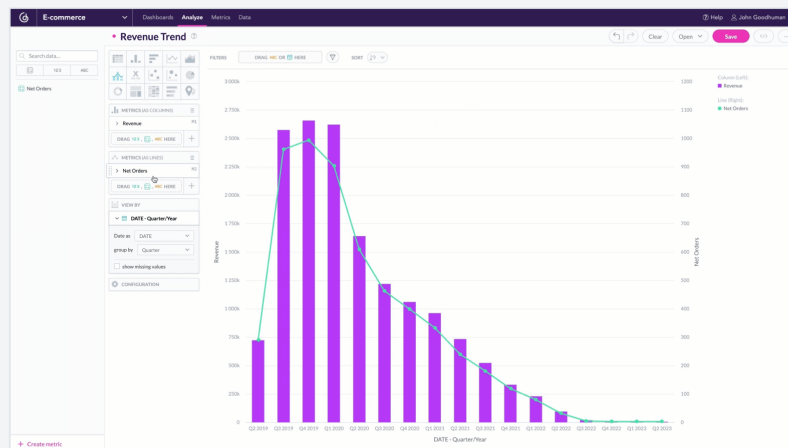
Analytics Engine

The GoodData [analytics engine](#) takes metrics and queries, and with the help of information coded in the LDM, translates these to SQL. Your database is queried directly through the analytics engine and the returned results are cached before being passed to the end user. The analytics engine can be utilized either through the API directly, or via the web interface.

Responsive Analytics UI

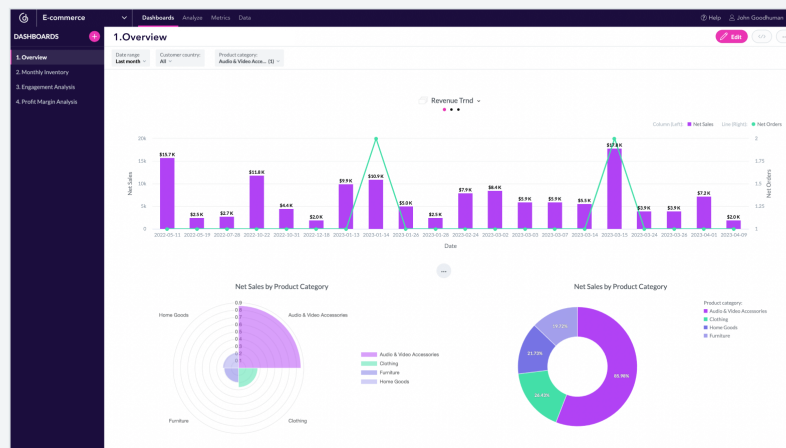
The GoodData analytics platform provides a full suite of tools for simplifying the delivery of analytics to a wide variety of users with different skills and needs. These tools employ responsive design patterns, so the visualizations, dashboards, and applications will dynamically adjust to a user's device. Each of these tools can be used for self-service experiences, integrated with applications, or used as a standalone tool with white labeling options.

Analytical Designer (AD)



[Analytical Designer](#) is a powerful and intuitive, visual drag-and-drop exploration tool for data discovery and the creation of visualizations (reports, graphs, charts, etc.). It provides the business-user recommendations on how to slice and view the data as they interact with the tool. Coupled with the intuitive drag-and-drop design, Analytical Designer makes discovering new data insights easy for any user, regardless of skill level. Users can save, share, export, and embed the visualizations or add them as building blocks for dashboards.

Dashboards (KD)



Dashboards are responsive and interactive storyboards. They are easy to edit and build, allowing any user, regardless of skill, to organize the visualizations that have been created with the Analytical Designer into interactive storyboards and share them with others. You can add any number of dashboards to your workspace. Optionally users can dive into data exploration without the need to ever leave the dashboards. Additionally, dashboards include support for various types of drilling, data exports, embedding, and other such interactions.

Customization

GoodData offers a wide range of customization options, from look and feel adjustments, localization and timezone, to supporting the development of completely customized analytical solutions.

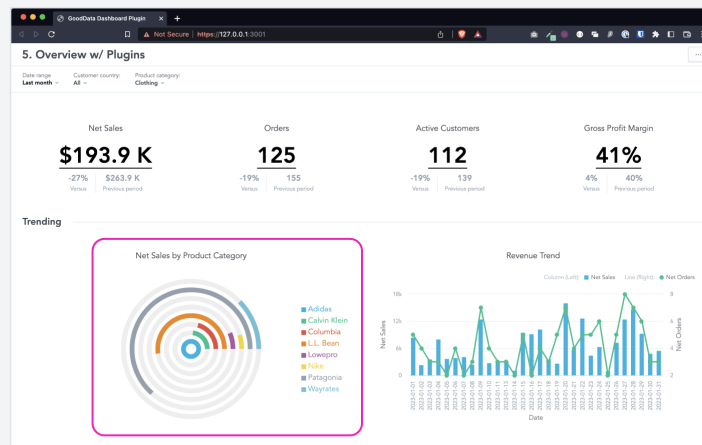
White-labeling & Theming

White-labeling allows you to remove and replace GoodData branded content with your own (this includes changing the hostname). **Theming** allows you to further customize the look and feel of Dashboards and Analytical Designer - such as changing the colors and fonts used.

Thanks to these features, GoodData can match the specific brand or design style you require, and seamlessly fit into your suite of applications – whether embedded or not.

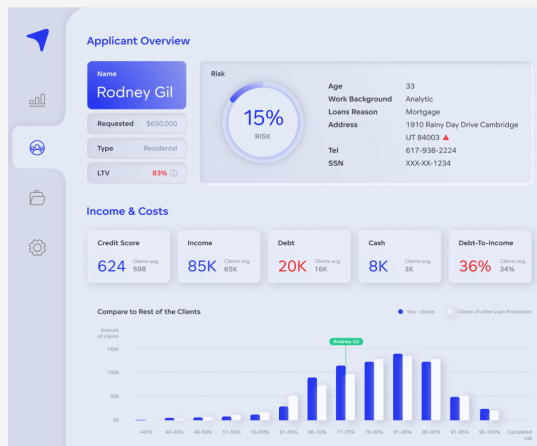
Dashboard Plugins

[Dashboard plugins](#) allow developers to create and integrate custom code into GoodData dashboards. With the plugins, you can customize and enhance the default dashboard experience available to the dashboard consumers. This allows you to add new types of visualizations to the dashboards, extend the functionality of the existing visualizations, or embed external code. You can also choose from various [dashboard plugins](#) in GoodData and request them from US.



Custom analytics applications/solutions

GoodData provides support for the development of custom applications on top of its analytical engine. Choose this route and you'll still benefit from GoodData's semantic layer and analytical engine, but you can develop a completely tailored analytics application. The two main routes our customers take are:



React SDK

[React SDK](#) is GoodData's React-based JavaScript library for building responsive analytical applications. You can use the SDK to build the Dashboard Plugins, embed visualizations and dashboards into your own application, and develop completely custom JavaScript applications on top of GoodData's semantic layer and analytical engine.

Embedding

GoodData allows you to easily [embed](#) visualizations, dashboards, and the Analytical Designer into your own web application. This enables your end users to access analytics without having to ever leave your application, increasing adoption and stickiness.

GoodData supports several different ways of embedding:

iFrame

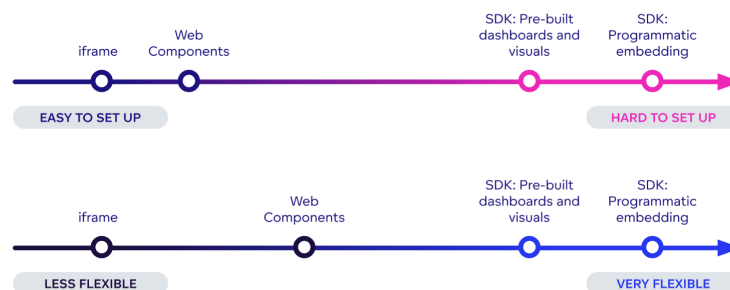
An [iframe](#) is used to display web content from one website inside another website. You can use iframes to embed your dashboards and Analytical Designer into your web applications. While iframes in general have some disadvantages (such as slower initial load), they often serve as the easiest and fastest way to achieve the desired result with minimal effort.

Web Components

GoodData provides a [Web Components](#) library that lets you easily integrate dashboards and individual visualizations into your web application. Web Components is a modern technology that doesn't suffer from the same disadvantages as iframes. Components built on the Web Component standards will work across modern browsers, and can be used with any JavaScript library or framework that works with HTML.

React SDK

GoodData's [React-based JavaScript library](#) is specifically designed to facilitate the easy embedding of existing GoodData visualizations and dashboards into your React applications, as well as to facilitate the creation of completely new components and visualizations that fit your exact use case. This option is the best fit for web applications written in React, but will work in other frameworks like Angular or Vue.js. Compared to iFrames and Web Components it may take a few more steps to set up, but offers the most customization and flexibility.



Multitenancy in GoodData

[Multitenancy](#) means that GoodData can manage an environment with many tenants, and each particular tenant can only access the entities and data that they are entitled to access. An unprivileged tenant has no access to those entities or data.

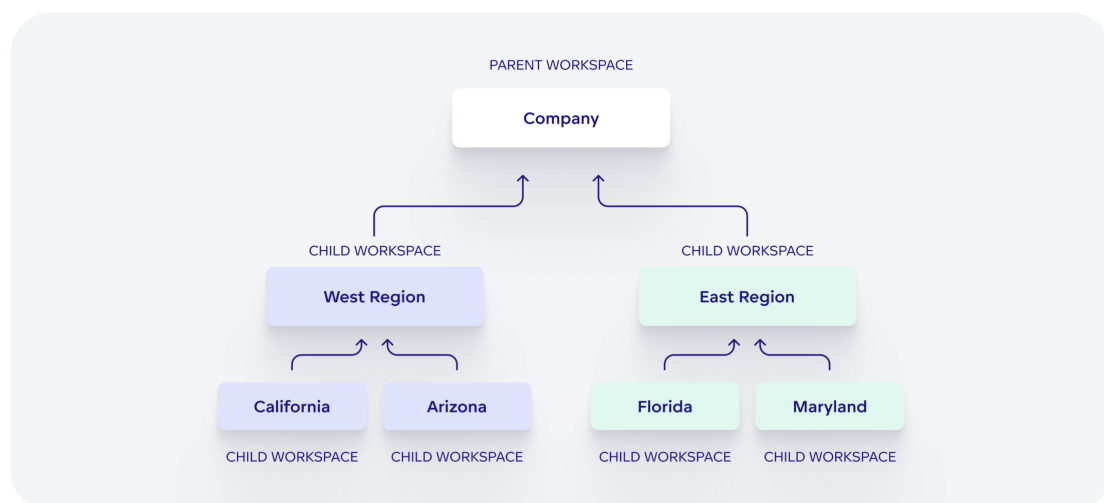
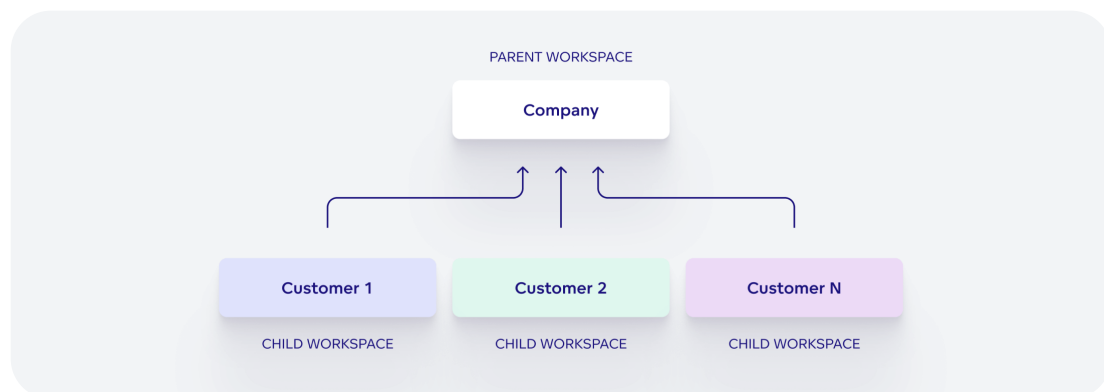
You can think of a tenant as:

- | Users and user groups outside your company who are related to your business (e.g., resellers, agents, franchise units, etc.)
- | Customers (e.g., subscribers or client companies)
- | Users inside your company (e.g., departments, global business units, or single users with specific needs)

In GoodData, multitenancy is managed by workspace hierarchy in which each tenant has their own workspace. Learn more about multitenancy [here](#).

Workspace Hierarchies

Workspace hierarchies consist of parent and child workspaces.



Each parent workspace serves as a template for its children. The children inherit all the entities from the parent workspace: the semantic layer, visualizations, and dashboards. Each child workspace belongs to a specific tenant, and the data accessible through the workspace are limited to data relevant to that tenant through the use of data filters – more on that later. The entities inherited from the parent workspace are available in ‘read only’ mode in the child workspaces. This allows for easy change management within the hierarchy. That said, while end users cannot modify the inherited entities, they can create their own visualizations and dashboards on top of the inherited ones.

Every change made in the parent workspace will appear in each child workspace, but changes made in different child workspaces will not affect the parent workspace or the siblings.

GoodData’s architecture has been designed with multitenancy at the forefront. This multitenant functionality allows multiple users to access the same data source and manage their separate workspace.

Workspace Data Filters

[Workspace data filters](#) allow you to limit the data available in child workspaces. By setting a data filter, you can define what subset of the data from a parent workspace will be available in its child workspaces. For example, a parent workspace may contain a visualization displaying the data from all company departments, but a child workspace will see only the sales department-related data in this visualization.

Child workspaces inherit the data filters from their parent workspace in the same way they inherit any other entity through the workspace hierarchy.

User Data Filters (Row-level security)

[User data filters](#) (also known as user data permissions or row-level security) allow you to restrict data that are available for specific users in specific workspaces.

By setting a data filter, you can define what subset of the data in a workspace will be available for individual users or user groups.

Single Sign-On

GoodData uses OAuth 2.0 and OpenID Connect (OIDC) to handle authentication. It allows users to log in using their existing credentials from a variety of identity providers (IdPs), such as Okta or Auth0.

GoodData integrates with the IdP to securely manage user authentication. This allows businesses to create a seamless user experience and embed insights into workflows through application integration

You can learn more about authentication and user management [here](#).

Analytics as Code

The concept of [Analytics as Code](#) is simple; we should treat analytics the same way as any other software. This means that analytics should be provisioned and managed using code and software development techniques, such as version control and continuous integration. With this approach, companies can significantly increase productivity and lower error rates.

GoodData gives you all the flexibility of a modern analytics platform. Thanks to the platform's extensive APIs and SDKs, you can switch between developing your analytics solution in an easy to use user interface and a programmatic approach as needed.

API-first

Traditionally, companies would first develop the product and then add [APIs](#) on top of it. In API-first, this mindset is reversed – APIs are built first and placed at the center of the product. GoodData concentrates on building reusable and easily accessible APIs that client applications can use and consume. The API is a core part of GoodData. All platform capabilities available through the user interface can also be invoked via the API.

Here are a few quick examples of what you can use GoodData's APIs for:

- | Integration of analytics with your own or a third party application
- | Automation of development, testing, and deployment of analytics using CI/CD pipelines
- | Export and import of declarative definitions for version control purposes
- | Programmatic consumption of data through the semantic layer – single source of metrics

GoodData provides a convenient OpenAPI definition and SDKs for accessing the platform functionality from multiple programming languages.

Declarative metadata

As mentioned above, GoodData allows you to manage analytics as any other code. In order for this to work, all the metadata from the platform (like workspaces, semantic layers, visualizations, dashboards) need to be exportable and importable. This is achieved through GoodData's declarative APIs. The metadata is exported in a declarative format that is easy to work with for both humans and machines. You can programmatically manipulate the metadata, use it to compose new objects, or version control it.

As a result, analytics becomes an easy-to-manage, reusable piece of code.

Python SDK

GoodData Python SDK provides a clean and convenient way to interact with the GoodData API in Python scripts and applications.

Python is a popular language for working with large amounts of data and data analytics. For this reason, we are actively developing this SDK to allow Python developers to integrate the GoodData analytical engine into their own applications as seamlessly as possible, or to automate their administrative workflow.

Python SDK allows you to script things that may otherwise be very tedious to do using just the GoodData user interface. You can find some examples of this in the next chapters.

```
sdk.catalog_data_source.create_or_update_data_source(  
    CatalogDataSourcePostgres(  
        id=data_source_id,  
        name=data_source_name,  
        db_specific_attributes=PostgresAttributes(  
            host=os.environ["POSTGRES_HOST"],  
            db_name=os.environ["POSTGRES_DBNAME"]  
        ),  
        schema=os.environ["POSTGRES_SCHEMA"],  
        credentials=BasicCredentials(  
            username=os.environ["POSTGRES_USER"],  
            password=os.environ["POSTGRES_PASSWORD"],  
        ),  
    ),  
)
```

Automate the provisioning

You can perform administration tasks such as managing users, permissions, and create new workspaces, as well as their hierarchies and data sources. With Python SDK you can write scripts that will let you easily create new workspaces, as well as manage existing ones.

```
# Reads visualizations from workspace
insights = sdk.insights.get_insights("123")

# Iterate through visualizations and check if they are valid
for insight in insights:
    try:
        sdk.tables.for_insight("123", insight)
    except Exception:
        print(f"Visualization {insight.title} is broken.")
```

Integrate into CI/CD pipelines

Integrate GoodData analytics into your continuous delivery practices by, for example, automatically deploying changes from your declarative workspace definition from GitHub to your production workspaces at an appropriate time in your production and delivery cycle.

```
pandas = GoodPandas(os.getenv('HOST'), os.getenv('TOKEN'))
df = pandas.data_frames(workspace_id="123")

campaign_spend = df.for_insight("campaign_spend")

# Now you have a dataframe with data from your visualization
# You can do linear regression, clustering, predictions, analysis, etc.
```

Create data pipelines

Export your data, leverage services like machine learning to transform your data, and import the data back into GoodData to visualize the results and gain insights. In the example below, we demonstrate GoodPandas, which can leverage machine learning practices.

You can learn more about GoodData Python SDK in [our documentation](#).

AI-enhanced features

GoodData leverages [Analytics as Code](#) as the foundation of its AI-powered analytics. In contrast to traditional drag-and-drop UI tools, our platform uniquely integrates [Large Language Models \(LLMs\)](#). These models can understand various programming languages and code structures, enabling the translation of natural language into structured commands.

GoodData utilizes LLMs to develop and test new models specifically for AI features, making use of its metrics store. As well as excelling at handling code-based structures, the semantic layer plays a crucial role in translating technical expressions and metadata into user-friendly terms. This ensures accessibility for non-technical users and facilitates the development of diverse AI features.

For first-hand experience of our capabilities, including AI features and other advanced functionalities, check out [GoodData Labs](#).

Deployment Options

GoodData is available either as software as a service, or as an application that you can deploy on your own server. Read on to learn more.

GoodData Cloud

If you are looking for a hosted, managed solution, GoodData Cloud is for you. With GoodData Cloud the analytics engine is hosted by GoodData in a public cloud, directly querying your data warehouse in real time.

Authenticated users can access GoodData Cloud from any modern web browser that has JavaScript enabled, or interact with the components directly through the API. GoodData cloud comes with two deployment options, based on your performance and security requirements.

Shared Deployment

Multiple customers share resources in shared deployment. Each customer has its own organization and is separated from the other customers on a metadata level. API calls and communication with the data source is achieved through public internet. TLS and IP whitelisting is available to secure communication. This option is the default one and suitable for most of the use cases.

Dedicated Deployment

Single Kubernetes cluster is dedicated to a single customer in this deployment model. This means that the solution can scale more flexibly. We can establish a private link between a dedicated GoodData Cloud cluster and your VPC. This option is suitable if you have strict security requirements or if you expect your solution to scale up/down dynamically.

You can learn more about GoodData Cloud [here](#) or start a [free trial](#).

GoodData.CN

GoodData Cloud Native (GoodData.CN) is a cloud native application ready for deployment on any Kubernetes cluster using a Helm Chart. You can utilize public cloud providers or host your own on-premise infrastructure.

Connecting your data in GoodData.CN follows the same process as with GoodData Cloud, except, as mentioned, you host the solution in your own public or private cloud as opposed to a GoodData-managed AWS cluster. If you are interested in GoodData.CN, please [contact us](#).

Enterprise Level Security and Governance

The GoodData analytics platform ensures the highest level of security by employing a proactive strategy that combines industry best practices and state-of-the-art technology. The platform employs a multi-layered approach to protect information, stay compliant with international standards and best practices, test and adopt new technology, and continuously monitor and improve applications, systems, and security processes — all while paying close attention to specific regulatory requirements in customer industries and locales.

For additional information refer to the [GoodData Security whitepaper](#).

Resources

The GoodData analytics platform offers robust and flexible implementation and analytical solutions to distribute valuable data insights to your end users. If you would like to test drive the platform's full capabilities, [register for a GoodData Cloud trial](#).

Find out more about the success of GoodData customers and sign up for webinars on the [GoodData website](#).

Read more about building analytical solutions designed for scale and performance in these [blog articles](#).

Learn more about the platform through our e-learning resources at [GoodData University](#).

Follow forums and knowledge-base resources through our [GoodData Community](#).

Developers can find more details about platform APIs and SDKs in the documentation or they can join us on [Slack](#).

Please contact us at info@gooddata.com if you have further questions or want help with assessing how the GoodData analytics platform addresses your analytical needs.