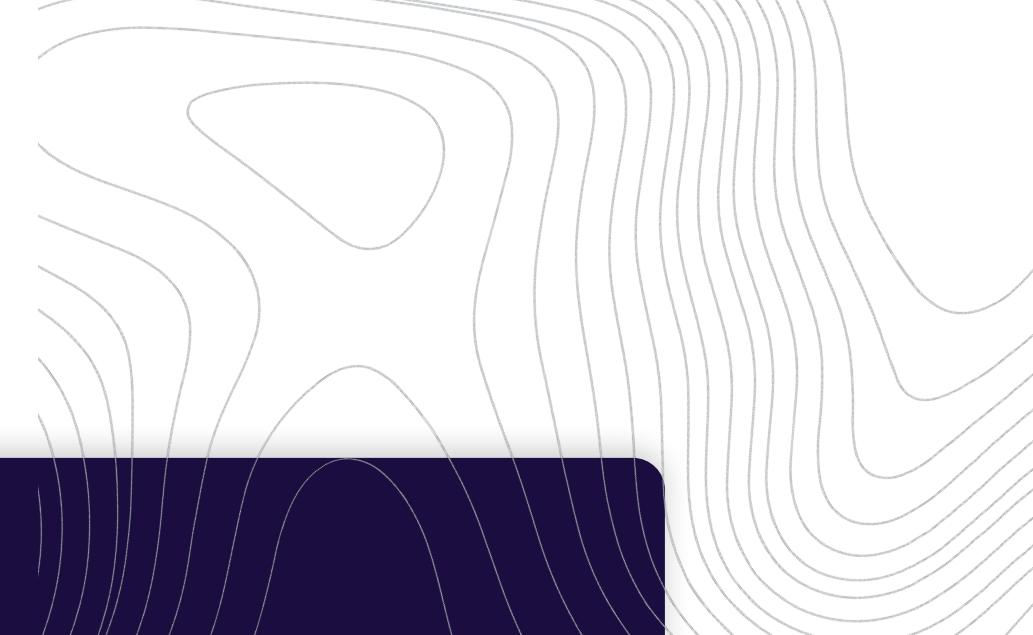




Scalability in BI and Analytics:

What, Why, and How



1

What is scalability in analytics?

Scalability is the ability of a system to adapt its performance and cost to changes in application and system processing demands. A platform or piece of software requires two main properties to successfully scale:

1. The capability to increase or decrease its capacity by adding or removing sources to adapt to the changing amount of data.
2. The capability to make this capacity available to those who need or want it.

A scalable business intelligence (BI) and analytics platform grows (and shrinks) with the needs of your business in terms of your data volume, your end user numbers, and how many business partners you have.

A scalable analytics platform enables this to happen **automatically, quickly, and securely**, while maintaining a separation between user groups – in terms of data, user privacy, and cyber security.

When we talk of analytics scalability, we are referring to the ability to use data to understand, solve, and meet a large variety of use cases and business goals. These use cases and goals come in many forms, so the analytics must be flexible enough to address different cases in different ways. This might include analytics **for internal BI or the use of analytics across other applications.**

2

Who can benefit from scalable analytics?

Scalable analytics solutions are particularly suitable for the following types of businesses:

- Companies that operate in the B2B market and provide analytics to their own customers (other companies or individual users of their products or services) but are unable to forecast the full extent of their customers' user bases.
- SaaS (Software-as-a-Service) companies that provide their customers with software which is designed to handle an increase in these customers and adapt to changes in demands.
- Data-driven companies that rely on data analysis technologies to organize their business and make future decisions, rather than relying on intuition, previous experience, or environmental influences.
- Companies with branches and stakeholders located around the world:
 - **Internal users:** When new employees join the organization, or when new teams are created (with more than one employee) based on different use cases, business areas, or organization departments – for example, finance, sales, marketing, product management, IT, etc.
 - **External users:** External parties such as partners who support the organization's business – for example, supply chain, logistics, investors, etc.
- Fast-growing companies with positive cash flow or earnings that are expanding faster than the economy as a whole.

3

How to achieve scalable analytics/scalability in analytics?

There are two main ways that companies can implement scalable analytics into their business (ideally a combination of both):

- **Scalable data storage:**

This requires a data storage solution that can handle a growing amount of data and support the business as it grows. This could involve using a third-party data warehouse, like cloud-based storage for example.

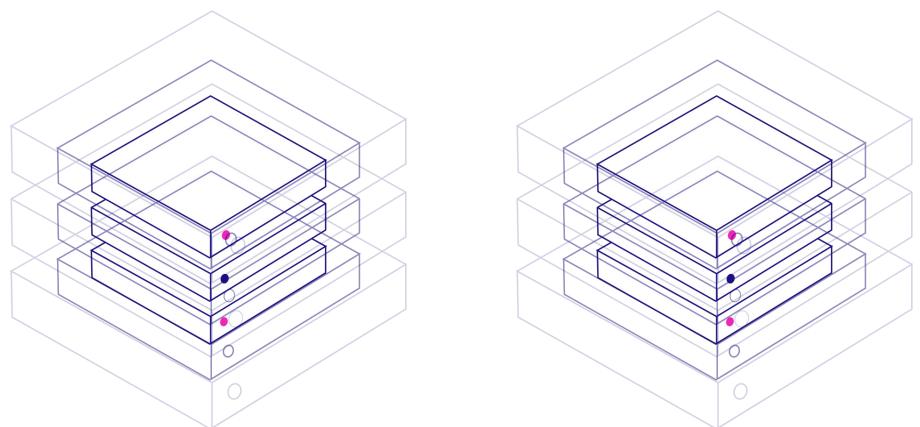
- **Scalable analytics platform:**

A scalable analytics platform allows the company to query large data amounts, effectively distribute analytics to a growing number of users, and integrate the analytics capabilities with various applications without affecting the analytics performance. This could include business applications, messaging platforms like Slack, or developer interfaces.

We can show how scalable analytics can be achieved in terms of data volume, users/user groups and use cases, and how this impacts the overall cost.

3.1 Data volume

From a data volume perspective, scalability guarantees that the infrastructure companies have purchased can withstand future data demands. Currently, we can divide companies based on the technology they use to store and manage their data — on-premise data storage or cloud data storage.

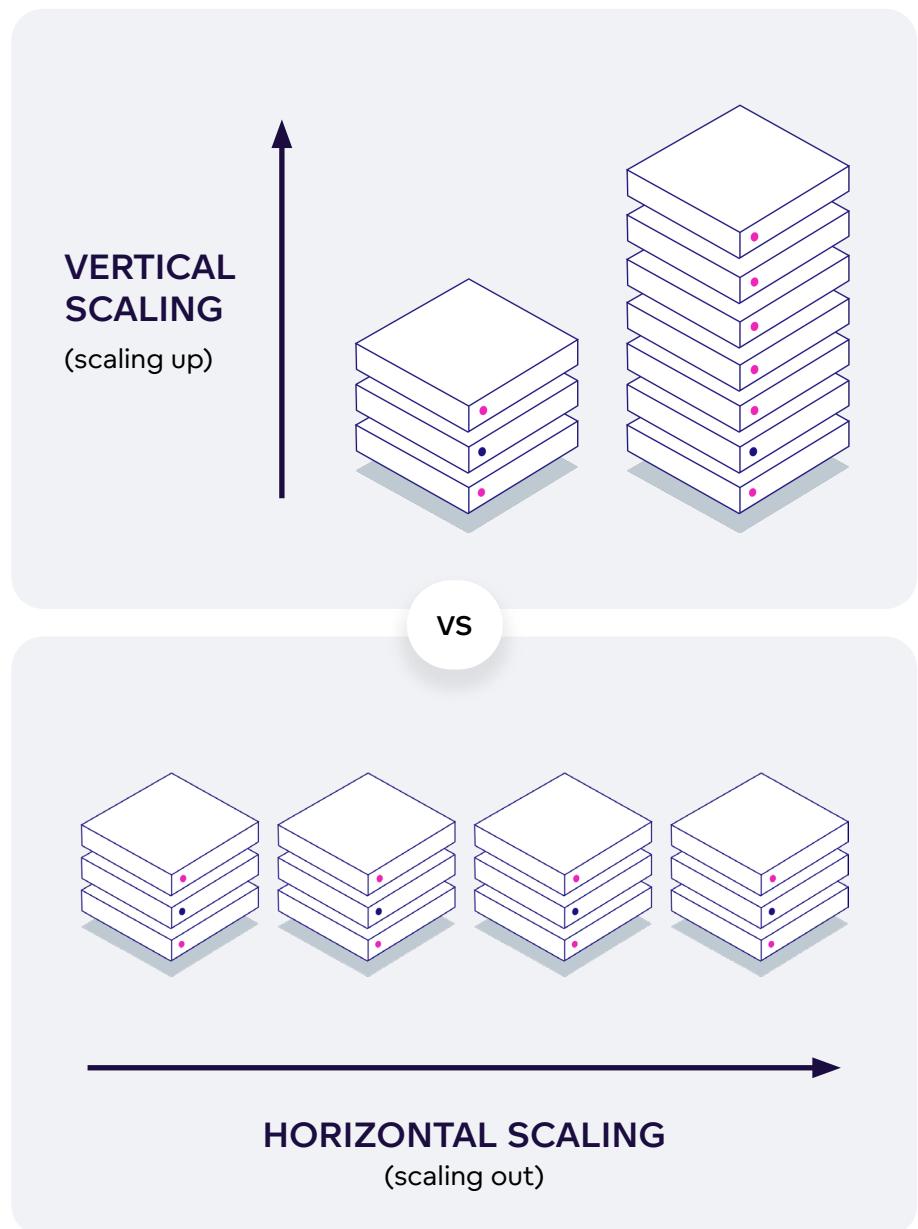


| Feature | On-premises data storage | Cloud data storage |
|------------------------|---|--|
| Cost | Initial cost for hardware and maintenance is high | Subscription-based, with costs based on usage and storage |
| Scalability | Vertical scaling/scaling up by upgrading/boosting current on-premises solution by adding more memory, storage, and processing power | Horizontal scaling/scaling out by adding more nodes in case of increased need for capacity |
| Responsibility | Companies are responsible for all hardware and their maintenance | The cloud provider handles the management and maintenance of the hardware |
| Performance | Can be affected by network congestion and other factors | Generally consistent and high |
| Security | Responsibility for security lies with individual companies The possibility of implementing specific security requirements | Security measures are handled by the provider |
| Accessibility | Enabling access mostly by users from the same location | Accessible from any device located anywhere via the internet |
| Limitations | There is a hardware limit and this cannot be endlessly expanded | The rental costs might be expensive as companies will have more licensed nodes |
| Recommended for | Small and medium-sized companies with stable data storage needs | Companies with rapidly growing data amounts |

On-premises storage provides control and security but may have limitations as data volume increases. Cloud storage allows for scalability and may be a good option for companies that outgrow their on-prem solution. Migrating from on-prem to cloud storage involves renting capacity, building a storage system, and transferring data.

Horizontal scaling is an appropriate way for businesses to simply add or remove resources – such as processing power or storage – to meet the demands of their workload. This form of scalability is also tied to [cloud-native](#) capability, and it can increase capacity automatically and seamlessly as the volume of data grows. As a result, it is well-suited for applications and workloads with rapid or unpredictable data volume increases.

Having scalable data storage is important in supporting large-scale data processing and analysis. But it is not the only factor. The analytical platform itself also needs to be designed to support large-scale data processing and analysis and should use techniques such as data partitioning and parallel processing to ensure that queries and analysis run efficiently and quickly.



Comparison of vertical and horizontal scaling

3.2 Users and use cases

When choosing a suitable analytical tool, it is necessary to consider whether it makes scaling possible. At some point, scalability usually becomes an important function that affects areas across the entire analytics solution. An analytics platform with the ability to scale should be able to:

- Unify data management and analytics throughout the organization or allow you to create your own solution. In this case, the analytics platform is expected to expand flexibly within the user base.
- Create a stand-alone tool that will work alongside other analytics tools and applications that are already in use in your organization.

It follows that the analytics platform should focus on scaling in a way that adequately responds to the user base and specific use cases. A focus on both of these factors is important if the analytics solution is to be technically feasible.

a. User base

Scalable analytics provides an analytics platform to individuals or user groups across organizations and their stakeholders. Modern analytical tools offer this solution, but some organizations still use traditional analytics built on a single-tenant architecture.

A **single-tenant analytics platform** refers to a single instance of the platform and its supporting infrastructure and database. This serves only one customer and lacks the ability to grow and account for additional users, which means that the analytics needs to be set up independently for each user on their device and each analytics must have a connected relevant data source.

The drawbacks of single-tenant analytics include:

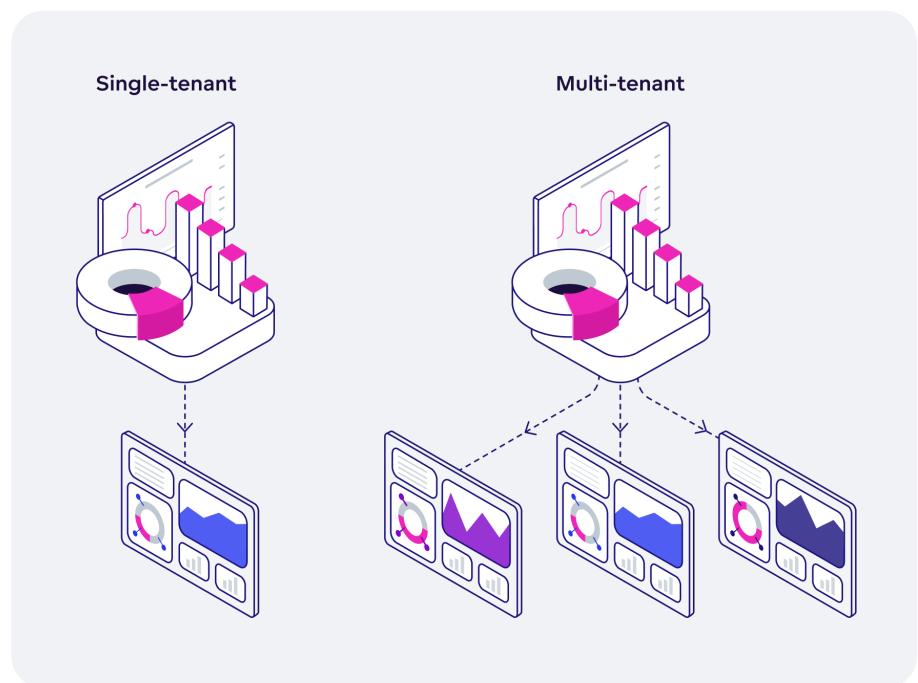
- The need to install analytics settings on each new user's device, which can be resource-intensive and costly.
- The inefficiency and cost of deploying new software versions for each user.
- The difficulty of managing all of the software versions and users as the user base grows.
- The inability to automatically deliver the same analytics setup to each user.
- The lack of suitability for organizations with global operations or external stakeholders.

A multi-tenant analytics platform allows companies to deliver analytics via a single platform that serves multiple tenants. Depending on the use case, tenants can be customers, departments, business units, or even single users with specific requirements.

Multitenancy enables companies to create an application just once by creating essential metrics, designing main dashboards, or providing other settings and deploying them to many tenants in separate [workspaces](#) for individual users and user groups. Multitenancy ensures that, within each tenant, users can only access the data they are authorized to use. Here, they can make changes within their workspace as they want.

A company can deliver analytics to their tenants — separate users or user groups – regardless of their geographic location. These tenants may be:

- Internal users of the company: Departments, global business units, or single users with specific needs.
- External users of the company: Vendors, resellers, agents, or franchise units.
- Company customers: Client companies who are paying for subscriptions to their products.



Single-tenant architecture vs. multi-tenant architecture

Multi-tenant analytics platforms are designed to be easily configurable without requiring changes to the underlying architecture, code, or data structure. This helps to save time and reduce costs.

b. Use cases

The best BI solutions are able to handle a variety of situations and use cases. A scalable analytics platform should allow a company to scale for different use cases and support its business goals. Some typical use case examples includes:

○ Internal BI and analytics

BI and analytics are often used by internal teams and stakeholders within a company. The goal of giving more internal stakeholders access to analytics is to enable them to make better decisions based on data analysis, facts, and insights, rather than relying on guesswork. Internal analytics can help a company increase profits and make more cohesive decisions by fostering a data-driven culture.

○ Integrate with different business/custom applications (embedded analytics)

Integrating BI and analytics with business applications can bring additional value to an organization. Instead of building a separate analytics service within their application, IT teams can embed the analytics interface directly into the application.

This allows organizations to easily incorporate analytics into the tools they use to perform their work.

○ User interface

A scalable analytics platform allows users and developers to access, modify, or obtain information about objects and functions through various interfaces, such as APIs or developer interfaces (Jupyter Notebook, Deepnote, etc.). This makes it easy for them to get the information they need.

○ Other communication tools

Scalable analytics may allow users to integrate analytics with communication tools like Slack and MS Teams. This enables them to easily obtain reports and information through simple queries without needing to log into the analytics platform.

Scalable analytics should be able to integrate with various applications to support a company's business goals, regardless of the specific use case.

3.3 Price

Before selecting a BI and analytics platform, it's important to consider your price expectations.

In a dynamic environment, the per-user pricing model can be expensive because it is tied to a specific number of users. It may not be cost-effective due to the difficulty in predicting the future user base.

A more flexible and scalable pricing model is important for a scalable analytics platform, as it allows the platform to adapt to changes in your business (such as fluctuations in the number of users) in a cost-effective way. The per-workspace pricing model is a good solution, as it charges a fixed fee for each workspace used, regardless of the number of users or activity within that workspace.

The per-workspace option can be a flexible and cost-effective pricing model for:

- Companies that have a large number of teams or departments and want to give each one its own dedicated workspace;
- Companies that operate in the B2B market and deliver analytics to their customers but cannot predict the future user base;
- Companies that have fewer users, or users who do not need to make queries and only require access to reports/dashboards to check business indicators/KPIs.

It is worth noting that a per-workspace pricing model may not be the most economical option for all companies, depending on their use case and the needs of the organization.

Per-workspace pricing can be flexible and cost-effective for companies with changing user numbers or a hard-to-predict future user base.

4

Scalability with the GoodData analytics platform

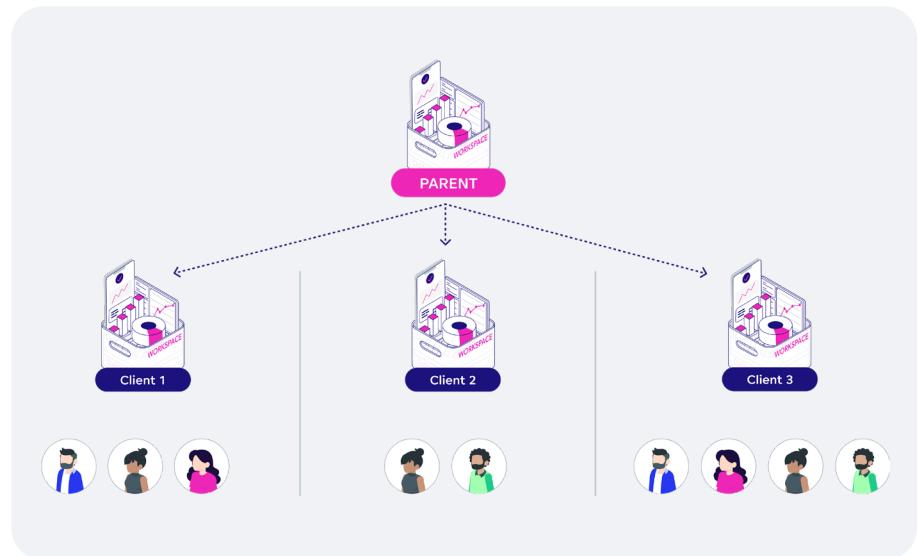
When deciding which scalable analytics platform is right for you, you might like to consider the GoodData platform. Here is a brief overview of how GoodData deals with scalability:

Data volume

GoodData has a distributed architecture that has been specifically designed to support the efficient processing and analysis of large amounts of data from multiple data sources. It uses techniques like data partitioning and parallel processing to quickly run queries and analyses on large datasets without degrading analytics performance.

Analytics platform

GoodData is a platform that helps organizations host and manage multiple tenants or customers on a single platform via different workspaces. It offers features like user and role management, data governance tools, and security and compliance features to help ensure the integrity and privacy of sensitive data. These features help organizations control access to data and resources at the tenant level and meet the needs of regulated companies.



Price

GoodData's workspace pricing model allows companies to buy a certain number of workspaces for their users — one workspace can be accessed by different numbers of users inside or outside of the company. It's also worth noting that the cost of a workspace may vary depending on the features and services that are included.

As a scalable analytics platform, GoodData also brings other benefits, including:

Automatic analytics delivery

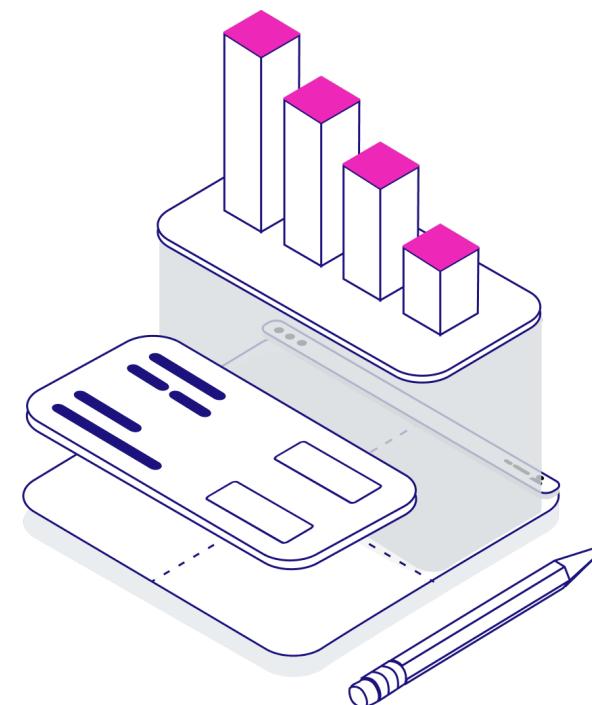
GoodData is a cloud-based analytics platform that allows organizations to easily and automatically deliver analytics to their users through customizable dashboards, metrics, reports, real-time data, and a robust API. Thanks to the parent-child relationship between workspaces, every change in the parent workspace is automatically displayed in the child workspaces. These features enable efficient access and use of analytics by users and user groups, enabling organizations to make data-driven decisions between different business units and improve their operations.

Self-service analytics

GoodData provides a self-service analytics platform that is designed to be easy to use. It allows all kinds of users to access and analyze data on their own without the participation of data teams. Thanks to [GoodData's semantic layer](#) and [MAQL](#), users are easily able to understand and utilize data, even if they don't have a background in data analysis or programming.

Embedded analytics

A range of tools and features enable organizations to easily embed analytics into their own applications and platforms, providing users with access to data and insights within their preferred tools. For more on this, check out our documentation on how to [embed & integrate](#).



5

Summary

The most important thing to remember is that scalability is a system's capability to effectively adapt to dynamically changing environments.

Other key takeaways from this e-book include:

- Scalability in analytics refers to a system's ability to adapt its performance and cost to changes in processing demands and data volume.
- Scalable analytics allows businesses to easily and quickly respond to new technologies and changes, as well as reduce the time and cost of maintaining analytics tools.
- Scalable analytics is useful for fast-growing companies, data-driven companies, companies with branches and stakeholders around the world, and companies operating in the B2B market.
- Scalability ensures that the infrastructure can handle a growing amount of data and support the business as it grows – for this the necessary data volume is required.
- Scalability allows for the support of a growing number of users, user groups, and use cases, integrating with various applications without affecting analytics performance.
- Scalable analytics uses the per-workspace pricing model, charging a fixed fee for each workspace rather than for each individual user.

6

Next steps with GoodData

If you have further questions you might find the following resources helpful:

- Learn about multitenancy in analytics in this blog post: [What is multi-tenant analytics](#)
- Check out this ebook to discover how to choose the best solution in relation to your user base size: [Architecture: Analytics for a Few vs. Analytics for Many](#)
- Download this to learn [How to Evaluate BI Tools to Choose the Best One](#)
- Learn how scalability can resolve bottlenecks here: [Anti-patterns for scaling analytics](#)

Want to try the GoodData platform for yourself? Do not hesitate to contact us for a [demo](#), or simply start a [free trial](#). Our team is ready to answer all your questions.