

# OLAP: The Key to Efficient, Scalable, Cost-Effective Data Lake Analytics





For decades, businesses have turned to online analytical process (OLAP) tools to help extract and analyze data. In fact, OLAP has been around so long that OLAP-based solutions may seem almost dated at this point.

In reality, however, OLAP remains as relevant and valuable as ever – at least for businesses that leverage modern OLAP tools, meaning those that can integrate with scalable, cloud-centric environments and resources. When deployed as an analytics solution for cloud-based data lakes, OLAP plays a central role in helping to optimize data analytics performance, manage data quality, minimize the cost of data operations, and, ultimately, derive the most value from large-scale data sets with the least effort.

This special report explains how OLAP has evolved to meet the data analytics needs of the cloud age and which specific benefits OLAP delivers as a solution for data lake analytics. It also offers tips on what to look for when evaluating modern, cloud-friendly OLAP solutions.

## The evolution of OLAP

Again, OLAP tools have long been a part of the solution stack that businesses use to understand their data. But today's OLAP solutions look nothing like the complicated, hard-to-scale tools that you might have deployed on-prem a decade or two ago to analyze a local database, for example.

Instead, we are living today in the age of what you might call "OLAP 2.0." Businesses can now benefit from a new generation of OLAP solutions that are far more powerful, flexible and scalable than their predecessors. They are capable of processing and analyzing data not just from local databases that are rigidly structured and limited in size but also from sprawling, loosely organized data lakes that run in the cloud.

The bottom line: If you think of OLAP as a legacy sort of analytics solution that has been supplanted in the cloud age by data warehousing or distributed processing systems, think again. While OLAP isn't the right type of analytics tool for every use case, it continues to offer immense value to modern business.



## Who needs OLAP for data lakes?

More specifically, virtually any business that stores data in a data lake – meaning a highly scalable, loosely structured repository that can house a wide variety of data – can benefit from modern OLAP platforms as a faster, simpler, and more cost-effective way of analyzing that data. Without cloud-friendly OLAP, it's difficult to process the data inside data lakes in an efficient, multi-dimensional way.



To be sure, it's not technically impossible to run analytics on a data lake without OLAP. You could, for example, convert your data lake to SQL databases, then run queries against each one. But this approach would be time-consuming, and it would significantly limit your ability to perform multi-dimensional queries by analyzing all of the data inside your data lake at once. Instead, you'd be restricted to querying individual databases.

Likewise, you could try to build a bespoke data analytics pipeline that depends on proprietary or custom tooling and complex query languages to analyze your data lake. The downside, in that case, is that you'd be stuck with a set of tools that are complicated and expensive to implement and maintain. You'd also become dependent on any proprietary tools you choose to deploy, making it difficult to update your tooling or switch vendors at a later date.

The easier and more obvious approach is to leverage a modern OLAP platform that lets you perform multi-dimensional SQL queries on a raw data lake – no matter how large and complex the data inside your data lake may be.

## The benefits of OLAP for data lake analytics

Now that we've discussed at a high level the main reasons to take advantage of OLAP as a data analytics solution for data lakes, let's walk through the key benefits of this approach in more detail.

### **Faster, simpler queries**

One of the most significant advantages of using OLAP for data lake analytics is that OLAP makes it possible to run queries directly against the data lake. You don't need to export, structure, or otherwise process the data in any way. You simply define your queries using a familiar language like SQL, then execute them.

All of the above adds up to queries that are faster to write and faster to run. By avoiding a data conversion process, OLAP minimizes the resources necessary to perform queries and maximizes the speed at which you get results. The fact that you can write queries using familiar syntax also speeds up the process because it means your team doesn't have to learn new or proprietary tooling to ask questions of your data lake.

## Governance

One of the inherent challenges in working with data inside a data lake is that the data is typically not well structured or organized. As a result, you can't apply the same types of access controls that you could in, say, a database, where it's easy to grant permissions to specific users for certain tables, columns, and so on.

With OLAP, however, you can enforce governance rules over your data without having to structure that data in a certain way. For example, you could enforce access control rules to manage who can do what with data inside the data lake.

Taking the governance features of OLAP even further, you could apply unified security policies to data inside a data lake. Such a policy would govern not only the data inside your data lake but also data in any downstream systems.

## Smart semantic layers

Another way that OLAP helps to address the challenges inherent in working with loosely structured data lakes is by making it possible to focus on the "semantic layer" of the data without worrying about underlying data.

In other words, modern OLAP tools can abstract the information inside data lakes in ways that align with data models or dimensions that users define. The data will then appear in this form to users, and they can run queries against the abstracted semantic layer regardless of what data exists in the underlying data lake.

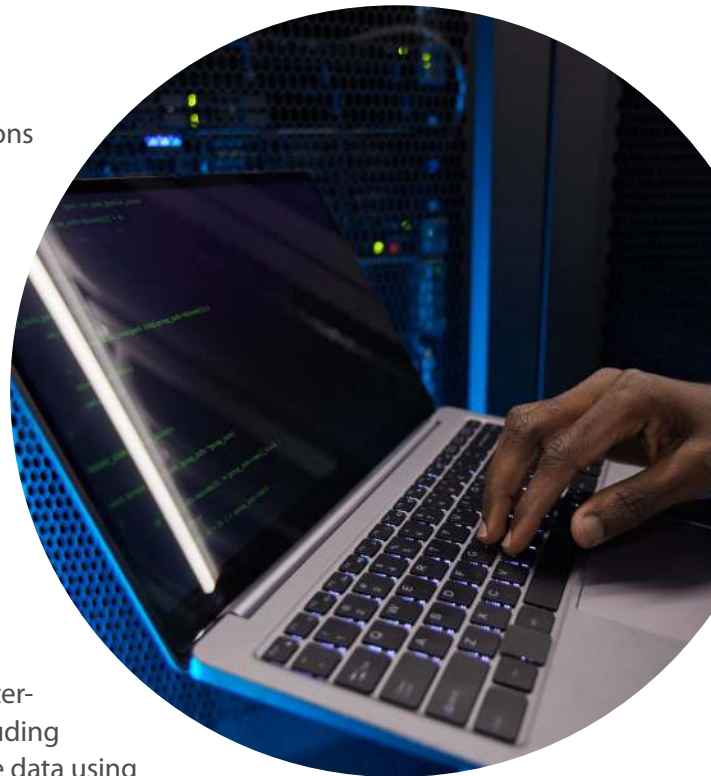
The result is a simpler, more cost-effective means of working with data, even in highly complex data lakes.

## Flexible, user-friendly analytics

Although SQL may be the standard query interface that your technical teams use to analyze information inside a data lake via an OLAP platform, it's likely that not everyone in your business knows how to write SQL queries.

That's not a problem when you use OLAP to analyze your data lake. Modern OLAP solutions let non-technical business users interact with data using a variety of business intelligence tools – including even Excel. There is no coding required, and users can access the data using a self-service analytics model without burdening their IT or data teams.

In this way, OLAP delivers the power of SQL-based data lake analytics for users who want it while simultaneously offering code-free, self-service integrations for other users who want to analyze data using a business intelligence tool of their choice.



## Minimal TCO

By enhancing data analytics performance, enabling self-service analytics, and ensuring that analytics operations can scale no matter how large your data lake becomes, OLAP minimizes the money you spend on data processing and analytics. Not only does OLAP avoid the expensive processing techniques of other types of storage systems, like in-memory analytics, but it also provides optimization features that can significantly reduce the cost of running queries on a data lake. For instance, you can take advantage of pre-computation, which lets you achieve stable query performance using a cost-effective "compute once, reuse many times" approach. This strategy enables enterprises to avoid the costs associated with the over-consumption of cloud computing resources.

In each of these respects, OLAP helps you to minimize the money you spend performing analytics on your data lake, but without requiring you to compromise on the depth or efficiency of your analytics operations.

## Conclusion

Far from being relegated to the shadows inhabited by "legacy" IT solutions, OLAP tools have evolved to meet the demands of large-scale, fast-moving data lakes in the cloud age. By taking advantage of OLAP solutions that allow anyone in your business to analyze information in a fast and simple way, businesses enjoy the power of multi-dimensional analytics for their data lakes without the cost and complexity of more expensive or time-consuming approaches that require data to be processed or migrated before it can be analyzed.

Kyligence helps organizations make the very most of OLAP-based data lake analytics. Kyligence not only supports multi-dimensional data models for data lakes of any size, but it also bundles in special features designed to optimize analytics operations. Examples include pre-computing functionality, which reduces the time required to complete queries, and an AI-augmented engine that automatically detects patterns from business queries and creates data marts based on them. Kyligence's unified SQL query interface ensures that you can execute queries easily across cloud object stores, cubes, indexes, and underlying data sources. And, because Kyligence scales automatically, it helps you analyze data cost-effectively whether you are dealing with mere gigabytes of information or many petabytes.



### About RTInsights

**RTInsights** is an independent, expert-driven web resource for senior business and IT enterprise professionals in vertical industries. We help our readers understand how they can transform their businesses to higher-value outcomes and new business models with AI, real-time analytics, and IoT. We provide clarity and direction amid the often-confusing array of approaches and vendor solutions. We provide our partners with a unique combination of services and deep domain expertise to improve their product marketing, lead generation, and thought leadership activity.



### About Kyligence

**Kyligence** was founded in 2016 by the original creators of Apache Kylin™, the leading open source OLAP for Big Data. Kyligence offers an Intelligent OLAP Platform to simplify multi-dimensional analytics for the cloud data lake. Its AI-augmented engine detects patterns from the most frequently asked business queries, builds governed data marts automatically, and brings metrics accountability to the data lake to optimize the data pipeline and avoid excessive numbers of tables. It provides a unified SQL interface between cloud object stores, cubes, indexes, and underlying data sources with a cost-based smart query router for business intelligence, ad-hoc analytics, and data services at petabyte scale.

Kyligence is trusted by global leaders in financial services, manufacturing, and retail industries, including UBS, China Construction Bank, China Merchants Bank, Pingan Bank, MetLife, Costa, and Appzen. With technology partnerships with Microsoft, Amazon, Tableau, and Huawei, Kyligence is on a mission to simplify and govern data lakes to be productive for critical business analytics and data services. Kyligence is dual headquartered in San Jose, CA, United States, and Shanghai, China, and is backed by leading investors including Redpoint Ventures, Cisco, Broadband Capital, Shunwei Capital, Eight Roads Ventures, Coatue Management, SPDB International, CICC, Gopher Assets, Guofang Capital, ASG, Jumbo Sheen Fund, and Puxin Capital.

Copyright © 2022 RTInsights. All rights reserved. All other trademarks are the property of their respective companies. The information contained in this publication has been obtained from sources believed to be reliable. NACG LLC and RTInsights disclaim all warranties as to the accuracy, completeness, or adequacy of such information and shall have no liability for errors, omissions or inadequacies in such information. The information expressed herein is subject to change without notice.