kyvos

WHITEPAPER

Universal Semantic Layer on GCP for Instant, Actionable Insights

Abstract

Key stakeholders and decision-makers rely on comprehensive, consistent and relevant data to strategize business objectives, plan operations and measure the results. However, with information flowing in from different environments, geographies and departments, it becomes difficult to trust data gathered within the business intelligence (BI) ecosystem. It is a rabbit hole of 'what-ifs' and 'what-should-be' since the gap between the 'data language' and 'business language' can sometimes be massive.

A semantic layer closes this gap with a consistent, current and unified view of business metadata across the enterprise for all users, irrespective of their expertise, skills and data proficiency. Gartner calls the semantic layer "a valuable enabler of self-service analytics," and it's spot-on.

The layer helps create a data-driven culture within an organization by allowing democratized, secure and self-service access to data. It simplifies both - the metadata existing within the data layers and business logic-related concepts outside this layer.

In this whitepaper, we'll explore how a universal semantic layer from Kyvos can hide the physical complexity of data from business users and provide actionable insights within seconds, irrespective of the data volumes or BI tools.

Table of Contents

Overview

A semantic layer is a layer of abstraction that provides a consistent way of interpreting data. It maps complex data into familiar business terms so that users across the enterprise can access the same source of truth with full confidence in its integrity.

Every analytics tool has been equipped with its own tightly coupled semantic layer since the concept was first adopted in the 1990s. Users have been walled in by the tool used by their organizational teams and challenged by siloed and limited data. Enterprise data management has only continued to get more complicated as data volumes started exploding beyond the scalable limits of most BI and analytics tools.

In the present context, data consolidation happens mostly in cloud data platforms like GCP, and business users rely on a specific tool for analytics. While modern BI tools solve the challenges of data trust, consistency and coherence to a great extent, they aren't enough when data grows in volume and complexity.

Traditionally, departments have relied on disparate analytics solutions to analyze the subset of organizational data which is of relevance to them. This approach can create unintended consequences, as business definitions and logic in each tool can lead to multiple versions of truth for the same data, especially when it's isolated in different locations.

Multiple BI is also a reality, with organizations using an average of 3.8 BI tools, according to 360Suite's BI Survey.

A unified view of data from different sources helps create curated and standardized models for consistent metrics across dashboards.

In addition, querying any amount of relational database requires a technical understanding of metadata and SQL to create a simplified business view of all measures, hierarchies and dimensions for in-depth analytics. Most business users do not possess these skills.

A semantic layer is the missing link between GCP and business users that masks the complexity of technology, standardizes and unifies enterprise data and introduces ease-of-use to complex business logic. A McKinsey study found that successful organizations make their data accessible across the board compared to companies where it lies in the hands of IT, data analysts or data engineers only.

Why Businesses Need a Universal Semantic Layer on GCP?

Managing data quality, multiple silos and numerous analytical tools—each with its own data definitions and models—can be challenging, especially when data volumes explode or if enterprise data exists in several locations.

Without a universal semantic layer in the analytics architecture, business users and IT teams draw differing insights, with no control over data access points and final interpretations. The challenges don't end here and include:

Data Trust Issues

Data scientists have been trying to bridge the gaps between trustworthy and untrustworthy data for years. Yet around 75% of business users don't trust enterprise data, and most of them won't make decisions based on this data. So, where does that leave an organization already grappling with too much information, used concurrently by numerous people, over multiple platforms?

When key decision-makers lack confidence in their data, delays occur, leading to a loss of opportunities. Although users can often conduct ad hoc reconciliations on any analytics tool, proving data quality is another ballgame. Inconsistent metrics may lead to differing interpretations and cognitive conflicts among teams.

Even with data democratization, without a semantic layer enabling truly self-service analytics, establishing complete trust in shared data and understanding the storage or relational representation of data can be significant impediments to the adoption of analytics. On the contrary, organizations using the semantic layers are 54% more comfortable with self-service analytics than others.

Multiple Versions of Business Logic

The data needs in an organization vary from one department to another, especially when teams use different analytics tools for analytical purposes. Let's imagine a scenario where a key stakeholder needs information about sales data for a particular region. This will require analysts to write codes for dashboard creation and data visualization while data engineers will create pipelines behind the scenes.

The process requires defining a specific set of business logic and assumptions for every rendition of the dashboards—decisions not always made consistently during the process. For example, buyers in Region A may be referred to as Active Customers in one dashboard, while buyers in Region B could be classified as Customers.

Now, if another manager needs a different dashboard using the same information and defines Active Customers as buyers from Region B, there will be discrepancies in data interpretations as the same data was referred to as Customers in the earlier report. Or one region could define a financial year as a calendar year, while another could use a different 12-month period. Different departments can even have slightly different calculations for something as standard as 'revenue'.

These multiple versions of truth created by different dashboards can make it difficult to make sense of any amount of data. Users can find it hard to identify a singular customer ID, region hierarchy or customer key. This can lead to no common data representation due to different interpretations of the same fields and different versions of the same truth. Business users need an approach where they can define and standardize concepts once and reuse them multiple times—within a tool and across several tools.

Semantic layers eliminate expensive and redundant rework across tools.

Lack of Version Control Leading to Fragmented Reporting

Data lineage and history are inherently significant to flawless query performance. Comparing the latest reports – say a month back - to those from six years ago will lead data users to nowhere, even if the definitions within these reports were the same. The result will be faulty comparisons unless they perceive the difference historically. This requires complete visibility into the version of a business measure, say net profit, from older reports and modifying business rules to create the same calculation for current reports.

Also, when data is decentralized, it can mask interdependencies of systems and fragment business definitions. Analysis on a larger scale can result in different or alternative definitions of common business terms in reports or charts. It comes back to the underlying problem with analytics tools. They work as a common playing field for different user personas - BI analysts, IT engineers and key decision-makers who are usually business teams with no analytical or even technical backgrounds.

Since each user group or persona has a distinctive understanding of business definitions, their interpretations can often be inconsistent with others or lack the precision required for predictive analytics. Mostly, teams end up with a tradeoff between intuitive results and dimensions they can slice and dice.

The solution is to fix the different versions and definitions of the data in different IT systems within the organization for complete version control via data preprocessing. Organizations need an additional layer even in their modern data stack to bind all the tools and deliver common definitions for different metrics or logics.

Compromised Analytics

The outcomes of analytics will be inaccurate if the information is incomplete and with inaccurate distinct values.

For example, two groups in an organization are assigned to sales analysis for a region with similar questions having slightly different contexts. Correlation results, in this case, varied since the sales team focused on Customers while the development team was more concerned with Product Range. Both results point to different problems, resulting in conflicted and compromised analytics.

When different groups prepare individual reports on different analytics tools without a semantic layer, each tool will create its own model and custom calculations, delivering different results for the same data.

Consolidated data representation and business definitions can prevent these disasters by unifying data, even if it's from multiple and complex sources. A data culture requires complete trust in data integrity and consistency, no matter who views any specific metric or definition and how.

Data Scale and Usage Constraints

Analyzing a constant stream of information by multiple users concurrently can take a toll on any analytics tool. And this doesn't end here; it goes further when data volumes increase, complexities grow, and data sources expand. Running interactive queries directly on GCP entails long, slow processes and highly complex code.

Moreover, a survey on LinkedIn says that 97% of information workers across organizations are non-technical people who find it difficult to tackle raw data and relational databases. Business users who possess skills to handle advanced calculations or complex data modeling can be a rare find. They still need instant data access to resolve various business issues. Throwing more analysts or BI experts at the issue won't help either.

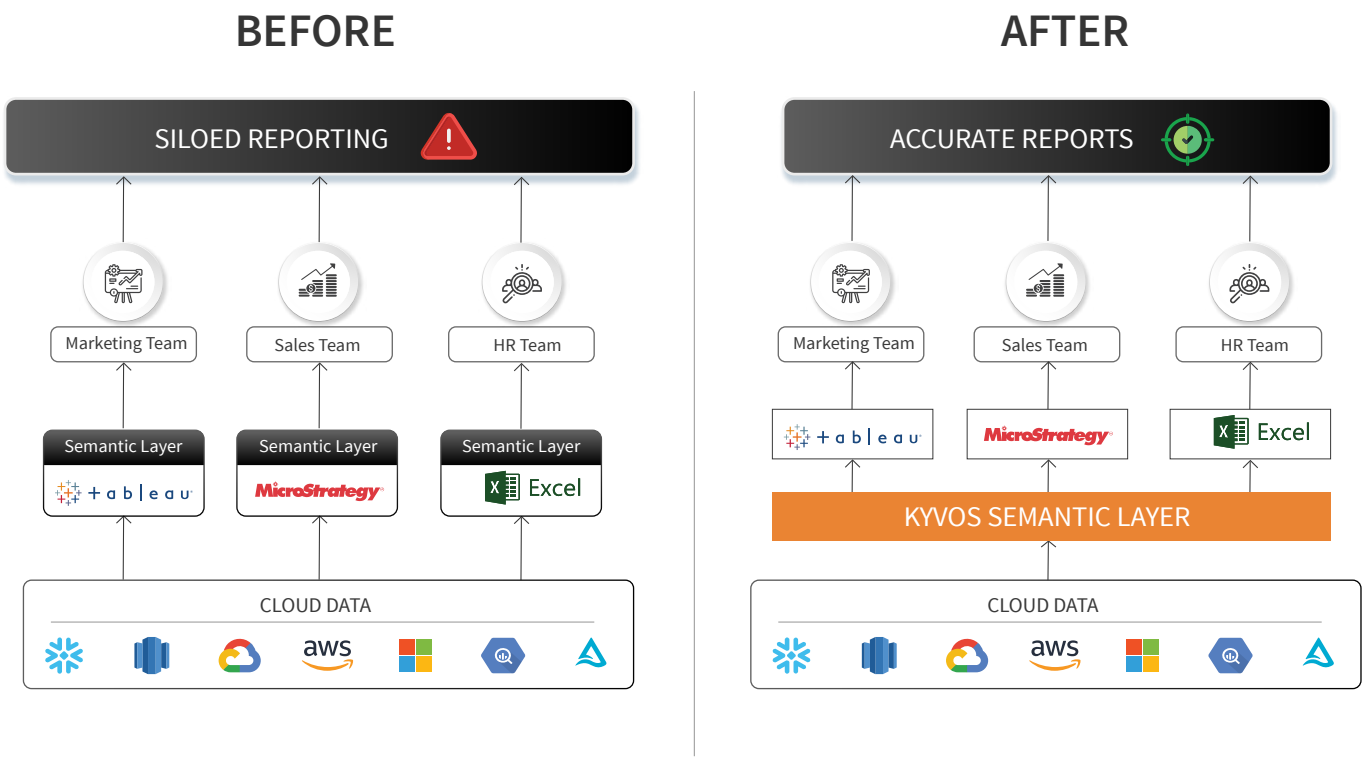
A robust analytical ecosystem can manage this future workload, especially when it grows to millions of cardinalities, billions of rows and hundreds of dimensions. While the traditional OLAP approach works well, it is mostly for structured data stored in semantic models. However, when fact tables (storing events or observations like sales related to a region) expand with the number of customers, sales or profits in a region, common queries can be preprocessed in tables by defining the most relevant fact tables.

But how can analysts decide which queries will be asked more often and how to update the preprocessed data over time for changing information? In these instances, the companies need a common layer that helps monitor common queries and look for frequent changes that need to be cached.

Universal Semantic Layer on GCP: A Valuable Approach to Simplifying Enterprise Analytics

A semantic layer on GCP adds business knowledge to the data layer in an analytics ecosystem to close the distance between data complexities and business users. It is an abstraction layer that represents analytics in a simplified, standardized business language or commonly understood terms, irrespective of the complexities of underlying data sources.

Enterprise data resides in the physical database, like a data lake or data warehouse. It is also the layer where the metadata, including data lineage, definitions, and profiling, is defined to better understand the data and ensure complete governance. A semantic layer is built over the raw data in an enriched and understandable data model by defining dataset relationships created by the code designers. Next comes the creation of semantic models with well-defined dimensions, hierarchies, attributes and measures.



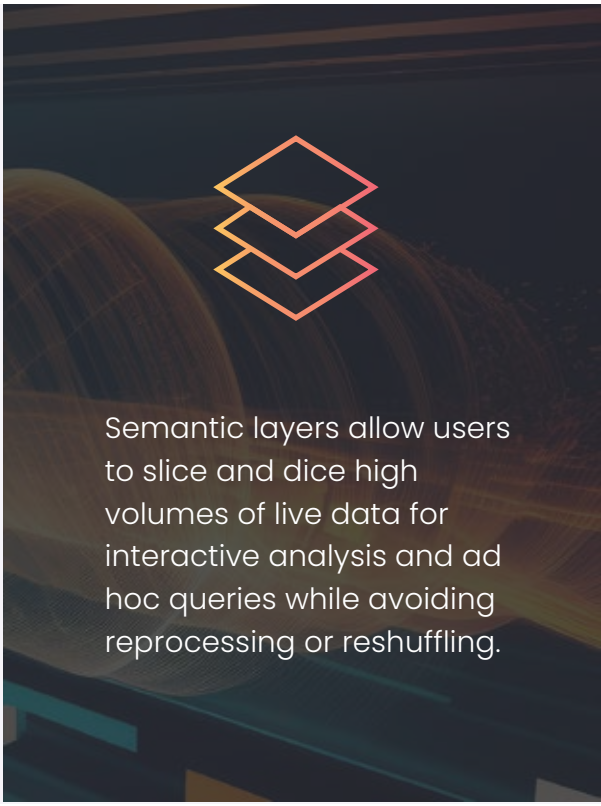
The end users can then view measures and dimensions available in their analytics tools to drag and drop them for instant querying and reporting. A semantic layer here works as a glue connecting all data, whereas the storage layer in the analytics architecture contains relational databases that only IT people can understand.

By standardizing business logic and creating the same view of the data for every analytic tool, semantic layers ensure that every user gets 100% accurate answers every single time.

Key components of a semantic layer can be:

- Columns with business nomenclature
- Semantic models with intelligently preprocessed data
- Pre-defined metrics with denormalized data structures

Here, it is important to understand that the layer won't compute queries per se as they are executed by compute engines built for this specific purpose. Consider it an 'analytics tool agnostic' layer that offers singular and similar references to data teams, business users and analysts alike.



How Does a Universal Semantic Layer on GCP Make Data Analytics Easier

From the business perspective, semantic layers hold much promise, especially when answering key questions leading to increased results, revenues and growth. Per Ventana Research, organizations using semantic models are 77% more satisfied with their data analytics.

The layer transforms metadata into a business view understood easily by anyone within the organization. A high-performant semantic layer can help modern organizations:

Establish an Enterprise-wide Single Source of Truth

The layer helps define a single source of truth, a version with a single representation of business terms that all stakeholders can use without a doubt. It interprets complex objects, long aggregate tables, join shortcuts and complex datasets for creating this version. The semantic model uses pre-built controls to create a version, where if two or more users ask similar queries, the results will be the same, faster and more accurate.

Another challenge of multiple data sources across an organization is that the process of integrating these sources is very long and complex. The integrated data always trails the current data, sometimes by as much as several months. In addition, multiple entities within an organization can create analytic products like reports and dashboards, based on data from different sources or recency. People do not trust data when multiple versions of metrics exist. The fear of the data being wrong then becomes a major impediment in making data-driven decisions.

Semantic layers solve the challenge of long and complex integration of multiple data sources across an organization. A single source of truth eliminates the resultant latency between the integrated data and the current data.

By establishing a single source of truth, semantic layers also address IT's challenge of balancing the need to truly democratize data and ensuring that business users get the right answers for their questions through self-service analytics, the next benefit we will look at.

Kyvos Universal Semantic Layer on GCP organizes, simplifies and accelerates data consumption. The platform synchronizes schemas and eliminates costly and redundant re-work across tools. It provides a consistent, standardized and trusted view of enterprise data.

In a similar use case, Kyvos' semantic layer built over an on-premise data platform helped a leading pharmacy chain preprocess 24 months of data and get instant insights on 315 billion records.

The layer transcends all restraints of data sources and formats. Whether the data is preprocessed or raw, data engineers can define business definitions and semantic models in a centralized way, irrespective of the tools deployed. The layer helps map any data format to a business-friendly interface that works for every user.

Get Business Abstraction to Enable Self-Service Analytics

The layer helps create a comprehensive view of data by uniformly maintaining business metrics, data access rules and calculation logic in one place – without the need of the expertise and experience required to correctly join multiple tables or deal with the resultant inconsistent data results. A flexible, self-service environment enables massive data to be parsed at the source and makes the information available as snapshots for quick and easy analysis.

Multiple business teams rely on data teams when the former needs immediate data access to make strategic business decisions, making the data teams a bottleneck in the analytical loop. A semantic layer can abstract complex terms to enable truly self-service analytics by eliminating this dependency.

The layer frees IT teams from ad hoc query resolutions while enabling business users without requisite technical expertise to answer questions like "How many new customers were added to Region 1 in Quarter 3 for Product X?" A semantic layer can process data models that remove SQL complexities and physical properties of the underlying data to deliver a unified and logical business view.

Semantic models are self-explanatory because data is organized by the way it's used by a business.

Semantic layers can mitigate the constraints like relying entirely on relational data models for business abstraction. While SQL is a common standard for enabling data access across data platforms, it might not suffice when data analysts or engineers prefer Excel or Power BI.

A universal semantic layer on GCP can decode all varied native dialects of each tool to deliver sub-second responses and consistent results for different queries. Multidimensional preprocessing for complex queries like semi-additive metrics or moving averages becomes easier by using this model.

Kyvos' next-gen semantic layer supports inbound languages like SQL, DAX, MDX and OLAP4J connectors so that every data consumer in an enterprise can query raw data using a tool of their choosing.

Multidimensional Analytics

A semantic layer creates the possibility of a multidimensional view by defining relational databases and preprocessing rules without any need for coding. Multidimensional analytics lets users deal with billions of accurate and approximate distinct count cardinalities, and slice and dice across hundreds of dimensions and measures.

The multidimensional data organization is the key differentiator that sets Kyvos' semantic layer apart from the rest. Kyvos' AI-based smart aggregation technology allows multidimensional analysis by creating semantic models on data that pre-process calculations on trillions of rows and enables MDX, DAX, SQL or OLAP4J querying that delivers instantaneous responses. As the entire data is preprocessed, all queries—standard or ad-hoc, old or new—are served equally fast.

In addition, combining schemas with varying levels of granularity without loss of details is not possible in a relational database without overcounting, which leads to the wrong answers. The Kyvos' semantic layer is one of the very few solutions that allow users to 'drill-across', addressing this challenge by getting 'distinct counts'.

The distinct count is the number of unique values in a column or expression, for instance, being able to accurately identify one unique customer through all the changes in their customer journey. Kyvos supports two types of distinct count: Approximate and Accurate. Using the Approximate count improves performance; whereas the Accurate count increases the size and processing time of the cube.

Kyvos delivers multidimensional analysis capabilities through its semantic layer that helps retailers solve the distinct count problem. It also facilitates calculating KPIs using both aggregated and non-aggregated measures – another unique feature of the Kyvos semantic layer.

The ultimate objective of a semantic layer is to deliver materialized data views by rendering OLAP semantic models for faster analytics. A fully featured platform like Kyvos enables the same with an automated optimization layer that can adjust according to skewing human behaviors.

"A universal semantic layer ensures live data access by connecting natively to GCP while delivering results at the speed of thought."

Pratik Jain,
Senior Technical Architect,
Kyvos Insights, Inc.

The layer helps improve query performance by caching the queries run earlier and optimizing the system with the least manual intervention. Since time is of the essence when making key business decisions, an advanced semantic layer can reduce delays in running analysis, even for billions of rows of data.

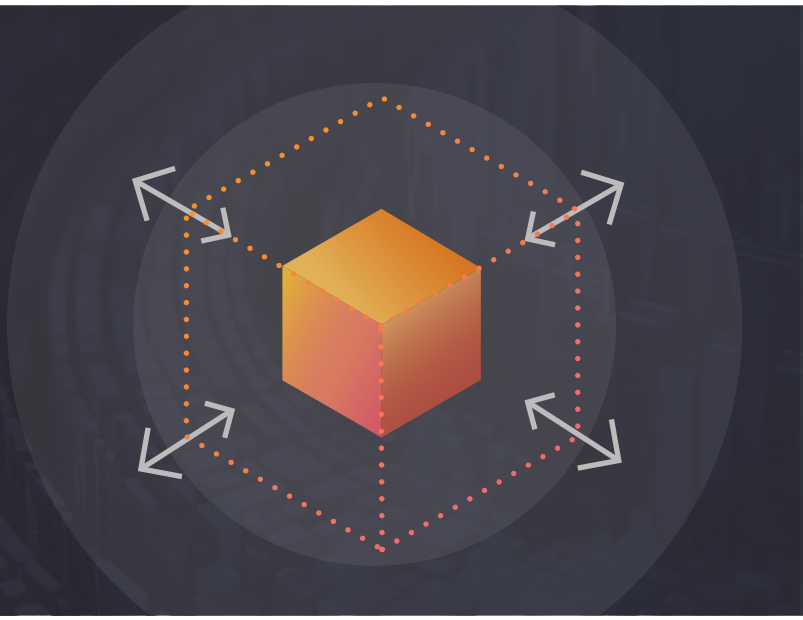
Achieve Scalability and Cost Advantages

A powerful semantic layer leaps over the caching layers in modern data stacks and provides interactive insights instantly. Data remains where it landed and there are no unrequired scans to increase the computing costs of GCP.

About 73% of enterprise data is never adequately used or analyzed, despite companies making substantial investments in data lakes or warehouses and the latest BI tools. Kyvos semantic layer increases data consumption by facilitating faster access to business logic irrespective of volumes or user concurrencies.

The repetitive queries are directed to preprocessed data and rewritten instead of being sent to data tables where they are read repeatedly, leading to much higher cloud consumption costs. So, it's safe to say that a semantic layer can reduce setup costs by optimizing data structures when business needs change.

Kyvos enables intelligent aggregation and caching with AI-powered self-tuning for sub-second queries on trillions of rows and high concurrency, with no performance degradation.



Reduce Data Engineering Efforts

Information workers rely heavily on raw data when extracting insights from respective data models. But this data holds minimal value when making critical business decisions.

Since data resides in multiple locations, relying on incompatible APIs isn't always feasible. Plus, breaking through silos of formats and locations can be challenging. A universal semantic layer allows data workers to blend data (raw and preprocessed data) across the platforms to create composite logical views on the server side. Once created, these shareable models serve every user in the organization.

Most users access data programmatically in sources; for example, Python or SQL notebooks, each with its capabilities. Since there are restrictions to what these interfaces can do, they may not even be supported by a specific analytics tool.

A universal semantic layer surpasses all these restrictions to create a unified view of data in business-friendly terms. It implements DRY (Don't Repeat Yourself) principles across analytics tools and enterprise data stack to easily track key business metrics.

Kyvos' smart semantic layer on GCP can turn raw data into consumable insights by adding a logical data model to the stack. This way, users can create a visual story that helps analyze even the most granular details in seconds.

A semantic layer can help reuse common logic even for high-level concepts. It defines the right building blocks for a code-free data model by adding virtualization features to any analytics architecture. This virtual layer is devoid of monolithic models as it abstracts elements that define business rules and the vernacular describing them. It joins multiple sources to future-proof massive investments against vendor lock-ins and allows users to query underlying data at speed.

Build Common Definitions for Different Use Cases

Multiple analytics tools mean copying the same business definitions repeatedly since the inherent semantic layers in these tools lack interoperability. A semantic layer can eliminate these issues with a user-friendly programmatic API (application programming interface) and metadata stored in an easy-to-understand format, irrespective of the analytics tool(s) used by multiple in-house teams. The layer decentralizes data for all tools with a standard framework to define measures, hierarchies and fact tables.

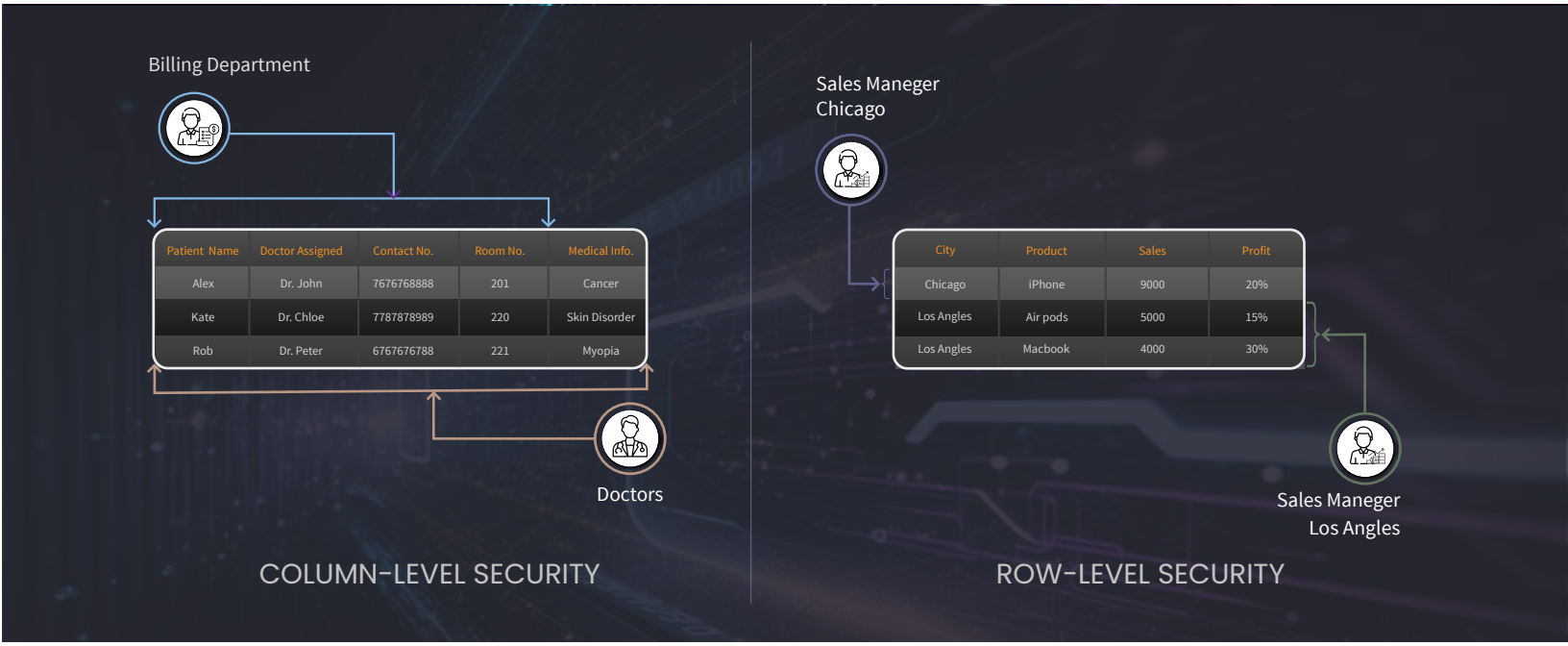
A truly universal semantic layer from Kyvos also ensures live connections to all programming interfaces, data sources and analytics tools for users across the board. From data engineers to business analysts, everyone wins with a semantic model offering complete consumption integration for connectivity to multiple tools and dialects.

"Kyvos enables trust in data. Experts build schemas inside Kyvos, but the business users running queries don't have to be one—yet their answers will always be correct regardless of their skill levels. They get the same answer to a question every time, thanks to Kyvos"

Leo Duncan,
Senior Solutions Architect,
Kyvos Insights, Inc.

Ensure Complete Data Security and Governance

Data lies at the core of a successful enterprise. As one of the most valuable business assets, it must be protected with solid guardrails at multiple levels from external threats as well as unauthorized access by those within the company's firewalls. However, adding a separate security layer to each level can be time-consuming and costly.



A universal semantic layer – which sits between analytics tools and GCP - is a perfect security mechanism for all analytical queries. Kyvos enables a multi-tiered architecture to ensure complete data protection across the stack. Users will get controlled access at group or individual levels for digging into granular details.

Authorization, authentication, encryption and role-based access controls flow throughout the applications. Users may be restricted from viewing specific measures, dimensions, or rows when accessing data. Centralized security rules apply to every user. As a result of similar integrations, more than 51% of organizations surveyed by Ventana Research found their data governance capabilities adequate after deploying a semantic layer.

Closing Thoughts

The universal semantic layer holds the key to a future where every data user speaks a common business language. The technology has already moved data from columns to concepts while creating a robust and scalable analytics infrastructure. Per a Gartner report, organizations sharing semantics and governance for better data sharing will outperform their peers in 2023-24. This looks plausible, considering the landslide changes in the data industry lately.

In the present VUCA (Volatile, Uncertain, Complex and Ambiguous) world of high data flow, old rules or legacy processes hold no value. Investing in too many platforms and tools will only grow an organization's technical debt and insight silos.

Kyvos' universal semantic layer works as the technical counterpart of the Swiss Army knife to cut through these barriers and help reach actionable insights by hiding data complexities, mapping and manipulation. It brings agility to business operations and ensures precise analysis of key metrics by any user.

About Kyvos

Kyvos is a modern, cloud-native, high-speed data analytics platform that enables sub-second querying on massive datasets. The platform's universal semantic layer democratizes data for all users across the enterprise, enabling self-serve analytics. Its AI-powered smart aggregation technology modernizes advanced analytics, while reducing the time and cost to extract insights. With Kyvos, instantly analyze data at any scale using the visualization tool and underlying cloud platform of your choice.

To learn more, request a demo now.

kyvos

FOR MORE INFORMATION:

www.kyvosinsights.com

info@kyvosinsights.com

@kyvosinsights

@kyvosinsights