

ATSCALE

Make *your data* speak the language of *your business*

The role of a semantic layer in scaling business intelligence
and enterprise AI

Table of Contents

4	It's Just a Matter of Semantics
6	Close the Language Gap between Data Consumers and Data Sources
9	Build Business-oriented Views of Enterprise Data in a Semantic Model
13	Let Data Consumers Interact with Data Using the Tools They Prefer
14	Leverage Query Virtualization to Simplify Access to Live Cloud Data
16	Deliver Speed of Thought with Automated Performance Optimization
18	Improve Analytics Governance and Reduce Risk
20	Leverage the Full Power of Modern Cloud Data Platforms
22	Top Use Cases for a Semantic Layer Platform
24	Take the Next Step

It's Just a Matter of Semantics

Data teams investing in centralized cloud data strategies have an opportunity to unleash data-driven innovation in their organizations led by business intelligence (BI) and data science teams.

THE KEY TO SEIZING THIS OPPORTUNITY IS BREAKING DOWN THE PHYSICAL, LOGICAL, AND SEMANTIC SILOS THAT NATURALLY EXIST IN ENTERPRISE DATA.

Centralizing data on cloud data platforms is a start to bridging these silos, but the challenge of understanding raw data structures and combining data from different sources still exists. Data is structured within enterprise applications in a way that minimizes redundancy and simplifies management. Naming conventions for tables and fields are oriented toward developers and architects, not business users. Even when raw data is transformed for use in a data warehouse, it remains difficult for business analysts and data scientists to understand how to properly interact with data.

The right semantic layer solution democratizes data access and unleashes an army of citizen analysts and data scientists.



A semantic layer abstracts away complexity while maintaining a live connection to raw data.

THE GOALS OF A SEMANITC LAYER ARE TO:

Let business analysts, business intelligence teams, and data scientists operate more independently and minimize time spent wrangling data.

Let different audiences create different views of data optimized for their uses while using a common set of definitions.

Create a single source for enterprise metrics, analysis dimensions, and model features — enabling centralized governance and ensuring consistency across use cases.

Support DataOps efficiency and harden data pipelines with a new control layer that integrates with broader enterprise data fabric.

The semantic layer is a foundational concept in enabling humans (and AI) to generate insight from data. Poorly implemented semantic layer strategies — or a lack of strategy — will result in metrics sprawl, shadow IT, inconsistencies, inefficiencies, and data mistrust. The right semantic layer solution democratizes data access and unleashes an army of citizen analysts and data scientists.



Close the Language Gap Between Raw Data and Data Consumers

The semantic layer sits between data consumers and data sources, presenting a business-oriented view that abstracts away the complexity of application data. Data consumers — whether BI analysts, data scientists, or decision makers — interact with views of data optimized for their use cases. A well-implemented semantic layer creates a single source of governed analytics that users can self-serve from with confidence.

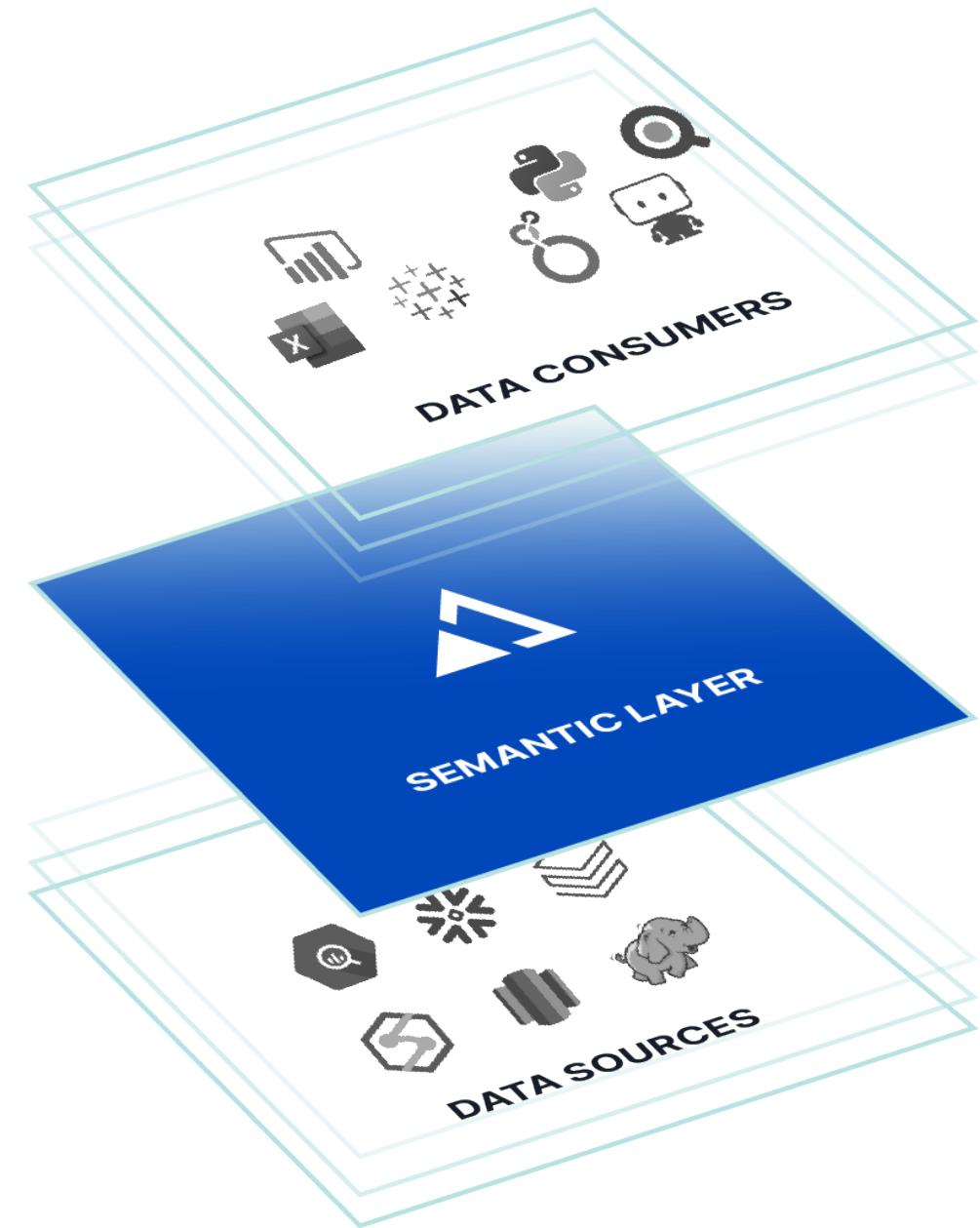


Figure 1: Linking Data to Data Consumers



While a semantic layer could theoretically be implemented within a BI platform or within a data warehouse, there are important limitations.

Implementing within a BI platform or a specialized analytics platform can provide tight integration with data visualization but precludes a shared semantic layer across teams that use different BI solutions.

Implementing within the logical definition of a cloud data warehouse is an option, but this limits the maintenance of the semantic layer to data architects.

According to Gartner, a semantic layer is a business representation of data that helps end users access data autonomously using common business terms. The semantic layer achieves this by mapping complex data into familiar business terms such as product, customer or revenue to offer a unified, consolidated view of data across the organization.

Source: Antelmi, Joseph, Demystifying Semantic Layers for Self-Service Analytics, Gartner, Inc., ID G00749457, September 7, 2021.





The Value of an Independent Semantic Layer

An independent semantic layer solution with open connectivity to consumption platforms and cloud data sources enables six key capabilities:

1

Fast and flexible creation of metrics and analysis dimensions for BI and feature sets for data science use cases;



2

Integration with the consumption layer tools that BI analysts and data scientists prefer;



3

Analytics that are connected to live cloud data through easy to manage data pipelines;



4

Analytics acceleration that delivers faster query performance and more efficient use of cloud resources;



5

Improved analytics governance and reduced risk through integration with security, access control and data catalog infrastructure;



6

Avoiding cloud vendor lock-in with freedom to choose from a broad range of cloud data platforms and the ability to manage migrations with minimal disruption to analytics users.



Build Business-oriented Views of Enterprise Data in a **Semantic Model**

A semantic layer solution should enable the creation of logical data models (i.e. a knowledge graph) that define a uniform set of metrics and analysis dimensions that are directly tied to raw data sources. The more intuitive and flexible the modeling solution, the more involvement business users and subject matter experts can have in the creation of new data assets.

WHAT ARE THE SEMANTICS OF DATA?

Semantic Models are the logical data models or knowledge graphs that define relationships of disparate data sources in terms understandable by business users. Metrics are quantitative – counts, sums, averages, etc... calculated at a level or aggregation defined by a specific cut. Dimensions are analysis categories organized in hierarchies. e.g. time, geography, products.

Features are the measurable pieces of data that can be used as variables or attributes within a data science model





PROJECTS QUERIES AGGREGATES SETTINGS SECURITY SUPPORT ORCHESTRATOR

Sales Insights - Snowflake MODEL CANVAS MODEL MATRIX MODEL DATA PREVIEW

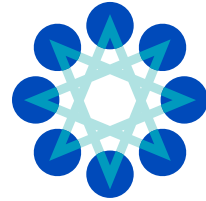
Internet Sales Model

Table	Column	Column Type	
factinternetsales	color	String	
	size	String	
	style	String	
	weight	String	
	productkey	Dec(38,0)	
	orderdatekey	Dec(38,0)	
	customerkey	Dec(38,0)	
	orderquantity	Dec(38,0)	
	unitprice	Double	
	salesamount	Double	
	taxamt	Double	
Calculated Tax	Double		
Relationships	salesordernumber	String	
	salesorderlinenumber	Dec(38,0)	
	productkey	Dec(38,0)	
	orderdatekey	Dec(38,0)	
	customerkey	Dec(38,0)	
	shipdatekey	Dec(38,0)	
	Order Dimension	Order Dimension	
		Currency	currencykey Decimal(38,0)
		Order	salesordernumber String, currencykey Decimal(38,0)
		Order Line Item	salesorderlinenumber Dec(38,0), salesordernumber String, currencykey Decimal(38,0)
		Customer Dimension	Customer Dimension
Customer Hierarchy			Customer Name customerkey Decimal(38,0)
Product Dimension		Product Dimension	
		Product Hierarchy	Product Line productline String
		Product Category	productline String, productsubcategorykey Decimal(38,0)
Date Dimension		Date Dimension	
		Date Month Hierarchy	Year year DateTime, Quarter quarter DateTime, Month month DateTime, Day datekey Decimal(38,0)
	Date Week Hierarchy	Year year DateTime, Week week DateTime, Day datekey Decimal(38,0)	
	Retail445	Reporting Year reporting_year DateTime, Reporting Half Year reporting_half_year DateTime, Reporting Quarter reporting_quarter DateTime, Reporting Month reporting_month DateTime, Reporting Week reporting_week DateTime, Reporting Day datekey Decimal(38,0)	
	CustomPP445	Custom Year reporting_year DateTime	

Figure 2: Semantic Modeling in AtScale

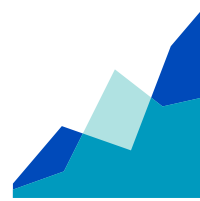


Semantic Modeling in AtScale



FLEXIBLE MODELING PARADIGMS

Business-oriented data modelers may prefer to interact with no-code drag and drop modeling environments. Developer-oriented modelers may prefer to create custom SQL modules to define more complicated calculated metrics or transforms. The best semantic layer implementation allows for flexibility based on these preferences.



ADVANCED METRIC/ FEATURE DESIGN

Users should be able to centrally define and manage a variety of metric calculations for consumption across all use cases, from basic quantitative measurements (e.g., sums and counts of fields) to more complicated ones (e.g., time-relative measures, compound-calculated measures). Data science users need to be able to engineer features including categoricals and booleans based on source data.



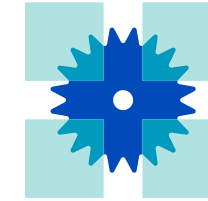
GROUNDING IN DIMENSIONAL MODELING BEST PRACTICES

Dimensional modeling lets data teams prepare datasets for analytics consumers that provide conformed dimensions and hierarchies to make it more efficient to navigate specific cuts of data.



OBJECT ORIENTED/ COMPOSABLE

Models and model components can be saved in a library for reuse. Rather than starting from scratch, modelers can build on existing dimensions, hierarchies, and complex metric calculations



MULTI-SOURCE

Modelers can blend data from multiple sources (e.g., disparate applications, third party data) with conformed dimensions (e.g., time, geography) with ease.

A semantic layer platform that integrates with BI tools, AI/ML/data science and application development platforms can deliver the benefits of self-service.



Let Data Consumers Interact with Data Using the Tools They Prefer

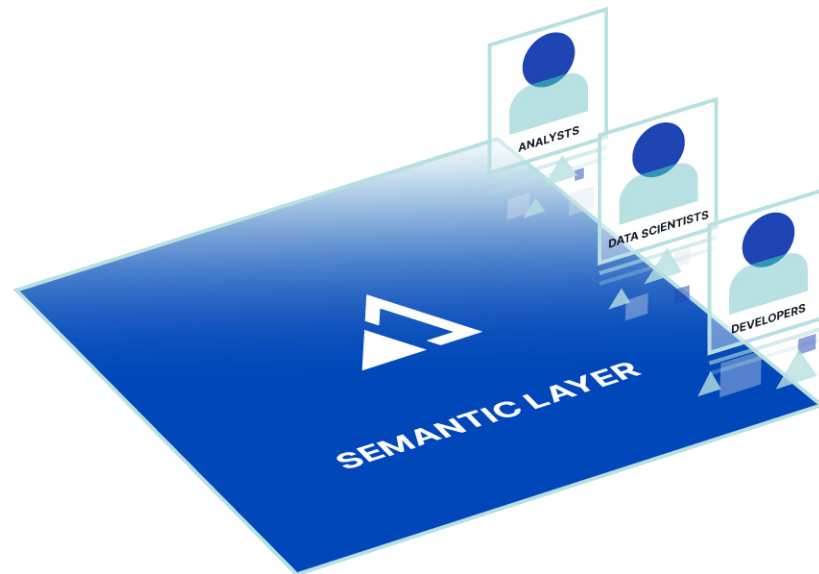


Figure 3: Open consumption layer / Integration with tools

OPEN/MULTIMODAL

Open integration with a full range of protocols for data consumption modes (e.g., REST, JDBC, ODBC, DAX, MDX and Python data) enables data consumers to leverage a full range of tools to consume data from the semantic layer.

NATIVE INTEGRATION / CLIENT OPTIMIZED

No client-side drivers are required when a semantic layer platform integrates with popular BI and data science tools with a SQL interface for Tableau, DAX interface for Power BI, MDX interface for Excel, LookML integration for Looker, and Python for AI/ML use cases.

MULTIDIMENSIONAL AND TABULAR

Multidimensional analysis tools like Power BI and Excel can fully integrate with a semantic layer that supports multidimensional protocols like DAX and MDX, including viewing of full data models.



How Query Virtualization Simplifies Analytics on Live Cloud Data

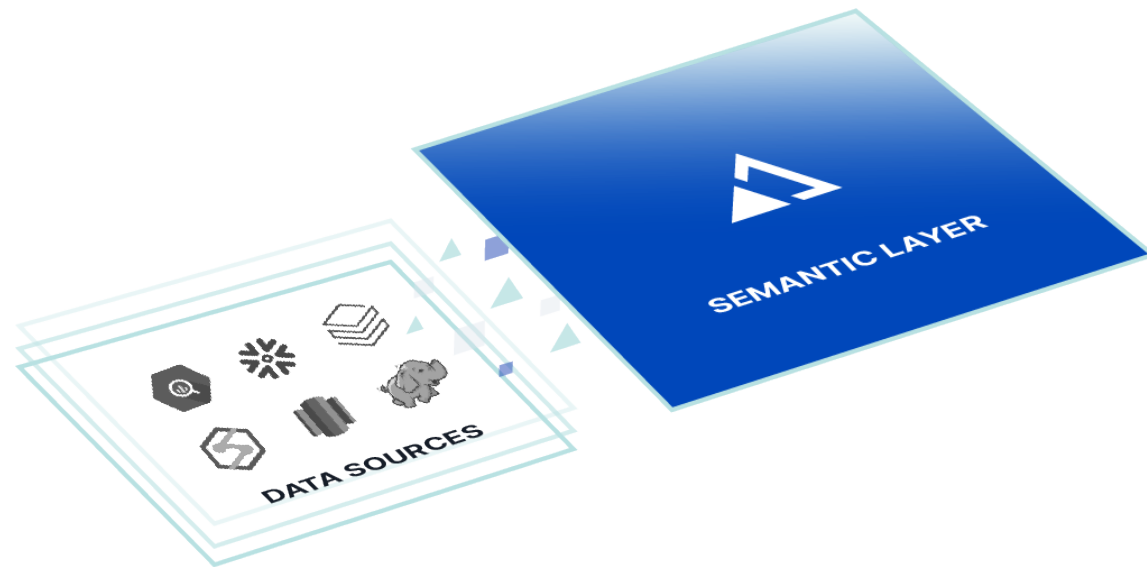


Figure 4: Query virtualization / Query optimization

QUERY VIRTUALIZATION

This capability gives users independence from creating and working with intermediate data, reducing complexity while increasing the recency and value of data. Data consumers can query semantic models tied to live cloud data, removing the need for data extracts, ETL or the need to write complicated SQL.

QUERY FEDERATION

Virtual data access keeps the data where it lives and allows blending of a reasonable volume of data across more than one data platform. Supports virtual data access without data movement.

QUERY TRANSLATION

Queries from the consumption layer, made against the semantic model, are translated in real time to physical queries in SQL dialect of data store. A graph planner can optimize SQL for performance and resource consumption in concert with automated performance optimization capabilities (e.g., substituting aggregate for base table based on previous queries).



The automation of aggregate definition and orchestration of data structure creation can optimize query performance and reduce cloud resource consumption.





Deliver Speed of Thought Analytics Performance for Cloud-Scale Data



INTELLIGENT AGGREGATE DESIGN

Performance-enhancing aggregations can be identified based on user query behavior and HLL-generated statistics. The more data and queries seen, the smarter, faster and more efficient queries become.

MATERIALIZE AGGREGATES

Creation, management and updating of aggregate tables can be automated and optimized according to platform-specific capabilities. DataOps teams do not need to manually create and maintain aggregate tables.

QUERY OPTIMIZATION

Real-time query optimization leverages aggregate selection using a graph-based query planner. As users query data from whatever tool they are using, the performance optimization engine intelligently routes queries to source data or aggregates.

Improve Analytics Governance and Reduce Risk

A semantic layer platform that integrates with data fabric elements, including security, access control, and data catalog infrastructure, supports greater discoverability and security policy enforcement — reducing risk and safely facilitating broader access.



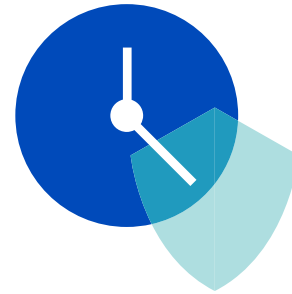
Figure 6: Data fabric integration / Analytics security





Access Control / Security

A semantic layer platform that understands an organization's directory integration with Windows Active Directory, identity security platforms and Online Analytical Processing (OLAP) technology can apply identity information in querying different data platforms (i.e., impersonation), including multiple cloud platforms. Centralized in a semantic layer platform, rules can be based consistently on directory policies, whether the data consumer is using Excel or Power BI, and respect security built into the data platform to accommodate admin users who may not have credentials.



Policies Enforced Just in (Query) Time

Policies may be applied to every single query at query time. There's no risk of creating an aggregate or in-memory copy that gets out of sync with a policy and leaves you vulnerable. Policies are enforced as you ask the questions.



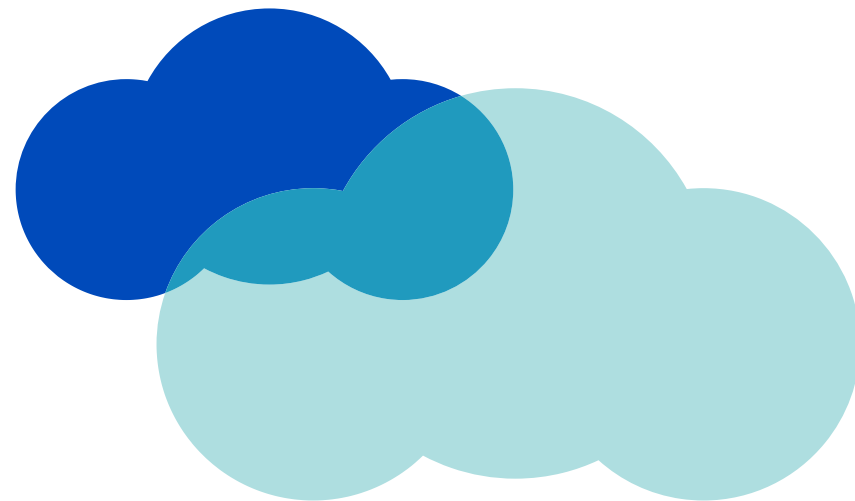
Data Catalog Integration

A semantic layer integrated with a data catalog can support data lineage and publishing of models in the catalog.





Leverage the Full Power of Modern Cloud Data Platforms



A semantic layer platform that integrates with those of cloud service providers ensures compatibility with a broad range of cloud data sources, enabling best-in-class platform selection with no cloud service vendor lock-in.

DATA LAYER INTEGRATION

A semantic layer platform with persistent connection to data platforms exposes structural metadata to the modeling environment during design and to raw data through the platform query engine.

DATA STORE AGNOSTIC

Integration with cloud data warehouse platforms (e.g. Snowflake, GBQ, Amazon Redshift, Azure Synapse SQL) and data lake platforms (e.g. Databricks, Cloudera Data Platform) gives BI and data science teams the freedom to choose from a broad range of cloud data platforms.

Top Use Cases for a Semantic Layer Platform

A semantic layer strategy is a critical component of a modern data and analytics program. There are three common use cases that start organizations on a path of re-evaluating their semantic layer strategy.

ANALYTICS ACCELERATION FOR LARGE CLOUD DATA SOURCES

A semantic layer platform is ideal for migrating away from legacy OLAP solutions, like Microsoft SSAS. Traditional OLAP tools rely on extracting data to create “cube” data structures that are time-consuming to manage, costly to store, and fundamentally disconnected from source data. As cloud data sources scale, it becomes impractical to maintain conventional OLAP for high-speed analytics. AtScale can deliver speed-of-thought analytics performance for live data connections to data platforms like Snowflake, Microsoft Azure Synapse, Amazon Redshift, Google BigQuery, and DataBricks.

ESTABLISH AN ENTERPRISE METRICS HUB

Using a semantic layer platform, organizations can democratize their data access by removing the obstacles between data sources and data consumers. Enabling self-service reduces the burden on highly-paid and hard-to-hire data scientists and engineers and puts more control in the hands of business analysts without introducing unnecessary risk. In addition, an enterprise metrics hub ensures a single source of truth and protects data from unauthorized access.

BRIDGE AI AND BI PROGRAMS

A semantic layer platform helps organizations bridge the gap between artificial intelligence and business intelligence. Data scientists can leverage the same infrastructure as BI teams for feature engineering and managing data pipelines. Further, model-generated insights like predictions can be published through the semantic layer for decision makers to access in the same tools and dashboards they use to work with historical data.





AtScale Can Help

AtScale is a semantic layer platform for data and analytics that helps organizations be more competitive by accelerating the rate of data-driven insights while reducing the cost, complexity, and risk of managing analytics infrastructure.

ATSCALE HELPS YOUR ORGANIZATION:

1

Promote data literacy and analytics self-service;



2

Accelerate shift to modern cloud infrastructure;



3

Reduce time spent with non-value-added data wrangling;



4

Get more value from data science / enterprise AI;



5

Eliminate lock in to single cloud platform or BI tool;



6

Achieve data analytics at the speed of thought.



Take the Next Step

About AtScale

AtScale enables smarter decision-making by accelerating the flow of data-driven insights. The company's semantic layer platform simplifies, accelerates, and extends business intelligence and data science capabilities for enterprise customers across all industries. With AtScale, customers are empowered to democratize data, implement self-service BI and build a more agile analytics infrastructure for better, more impactful decision making. For more information, please visit www.atscale.com and follow us on LinkedIn, Twitter or Facebook.

[REQUEST A DEMO](#) →

[CONTACT US](#) →

[LEARN MORE](#) →

Read [The Practical Guide to Using a Semantic Layer](#)

Download [The Complete Buyers Guide to a Semantic Layer](#)

Review [How does using a semantic layer impact cloud data warehouse performance? \(benchmark reports\)](#)

Get [advice from fellow data & analytics leaders](#) on how to scale smarter data-driven decision-making

[Stay current](#) on analytics strategies with short [articles](#) and [webinars](#) on topics in analytics strategies

