ATSCALE

Technical Overview

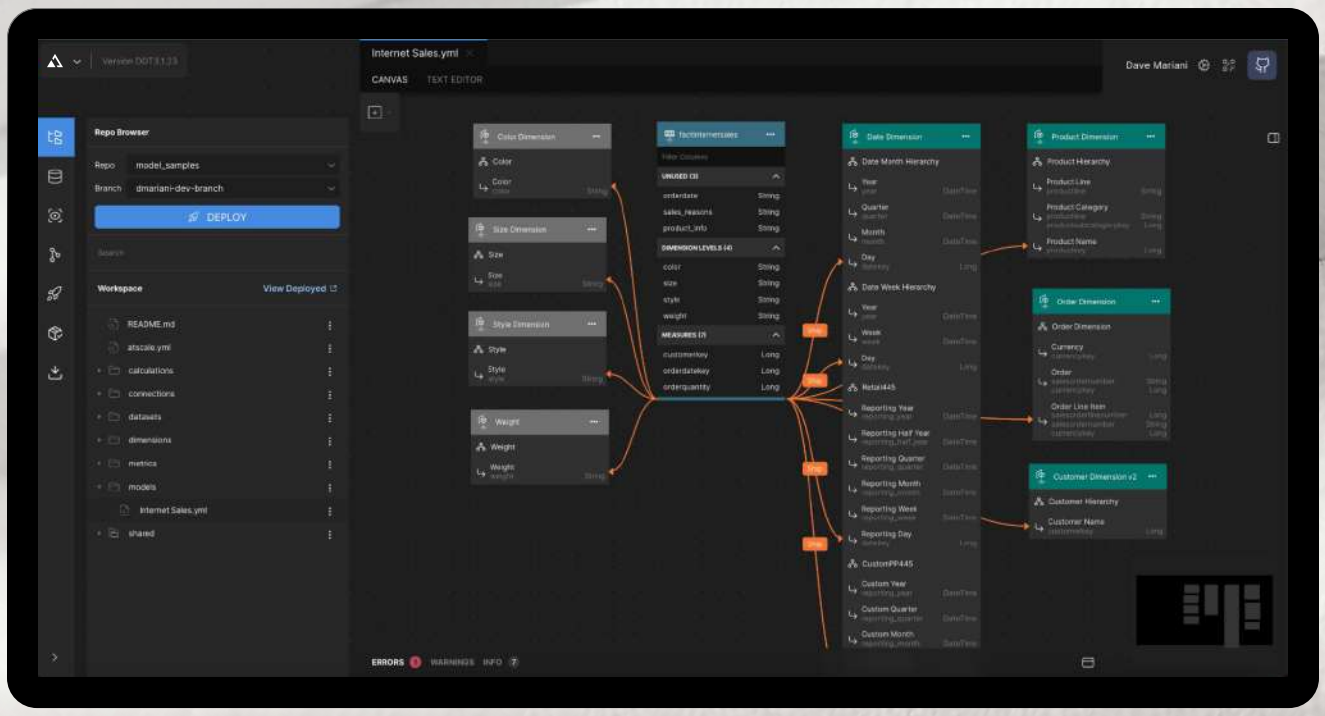


Table of Contents

Why AtScale?

AtScale offers a modern approach to business intelligence and analytics in the cloud. AtScale's Semantic Layer platform enables analysts to perform sub-second, multidimensional analysis with popular BI and AI tools. Enterprises rely on AtScale to overcome data and analytics challenges, including accelerating data-driven decisions at scale, creating one compliant view of business metrics and definitions, controlling the complexity and costs of analytics, and reducing the risk of analytics.

AtScale helps enterprises:

- **Seamlessly migrate to the cloud.** Enterprises can avoid business disruption and port analytical workloads with minimal disruptions to end users.
- **Simplify the analytics infrastructure.** Enterprises can use the best tool and platform for the job without moving data or adding new data stores.
- **Modernize and future-proof the analytics stack.** Enterprises can use data lakes and cloud data warehouses while preparing for future platforms.
- **Secure and govern data in one place.** With a live, governed connection to all data in a virtual model, enterprises can eliminate data copies and their associated security risks.
- **Turbocharge analytics and machine learning initiatives.** Enterprises can instantly integrate new data sources without manual engineering because AtScale delivers a single, super-fast, business-friendly semantic interface for all data.
- **See all data in a single, unified view**, no matter where it is stored or how it is formatted.
- **Conduct interactive and multidimensional analyses** using business users' preferred BI tools, such as Excel, Power BI, Tableau, or another tool.
- **Get consistent answers across departments and business units** via AtScale's Semantic Layer, standardizing queries regardless of BI tool or query language.

Where Does AtScale Fit in the Analytics Stack?

The AtScale Semantic Layer sits between your analytics consumption tools and your data platforms. By abstracting away the physical form and location of data, the AtScale Semantic Layer makes data stored in data lakes or data warehouses accessible with the same interface. Integration with enterprise data catalogs makes AtScale models discoverable and metadata shared seamlessly.

INSIGHT GRAVITY

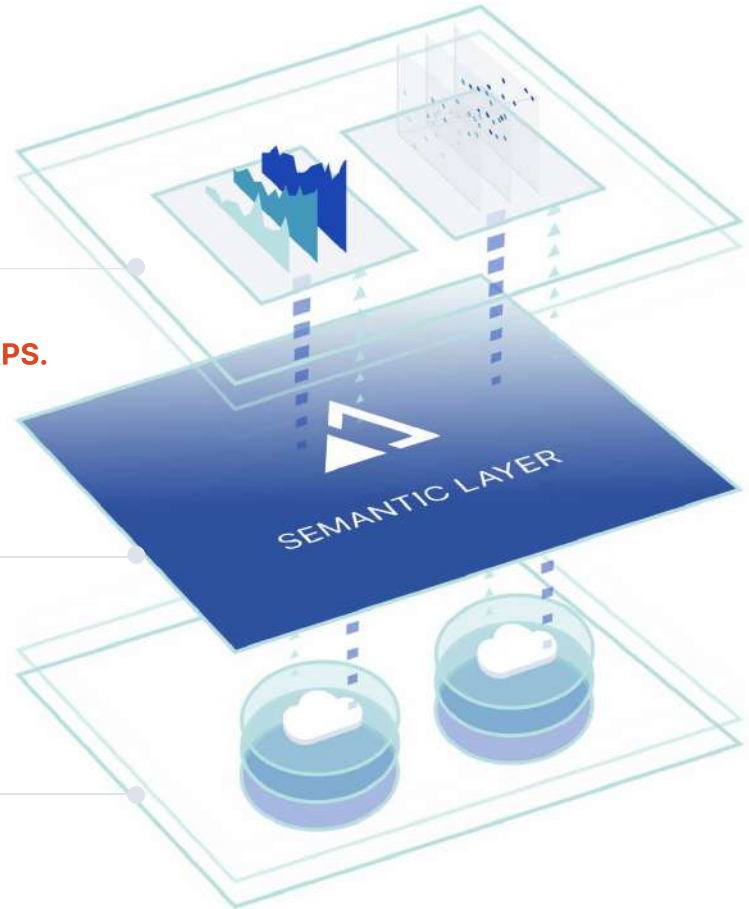
Continues to diversify across a broader spectrum of analytics approaches

COMPLEXITY. INEFFICIENCY. SILOS. SKILL GAPS.

The complexity of managing knowledge related to business context is a fundamental obstacle

DATA GRAVITY

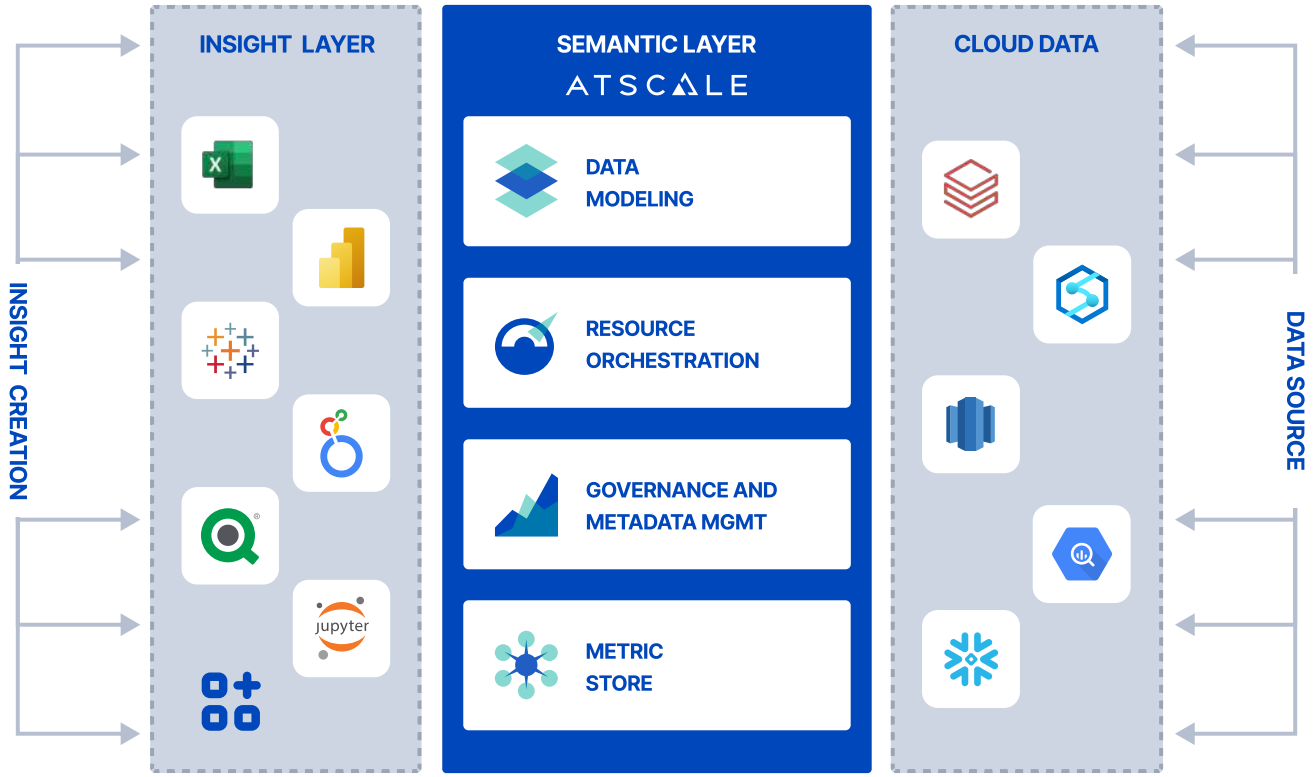
Continues to shift toward centralized cloud data platforms



AtScale System Component Overview

AtScale provides a single, secured, and governed workspace for distributed data. The AtScale Semantic Layer platform behaves like a logical data warehouse. The AtScale service intercepts client queries, translates logical queries into physical queries, and passes those queries onto the underlying physical data warehouse or data lake for execution. As end users interact with the data in the AtScale model, AtScale automatically creates or modifies aggregate tables to optimize performance and manage costs. AtScale will create aggregates (think materialized views) on the source data platform and determine the optimal location to store those aggregates in a federated query scenario. AtScale's automated tuning functionality works consistently regardless of the underlying data platform (data warehouse or data lake).

The combination of AtScale's semantic model, data virtualization, performance optimization, and analytics governance powers business intelligence (BI), artificial intelligence (AI), and machine learning (ML) initiatives resulting in faster, more accurate business decisions at scale.



The following sections will describe each component's inner workings and benefits.

Metrics Store

AtScale speaks the languages of your analytics applications, whether business intelligence (BI) tools, AI/ML platforms, or custom applications. AtScale requires no custom client-side software installations, so anyone using Excel, Power BI, Looker, or Tableau can connect to AtScale and run queries immediately.

Unlike other semantic layer platforms that offer only a single inbound query interface, AtScale offers a wide variety of interfaces optimized for each tool, so consumers with live connections to data platforms like Snowflake, Databricks, and BigQuery will not experience a degraded experience without data extracts, cube building, or imports.

Benefits include:

1. Support for multiple inbound interfaces, including SQL, MDX, DAX, REST, Python.
2. Support for Business Intelligence and AI/ML tools, as well as custom applications using REST, JDBC, ODBC, or XMLA interfaces.
3. XMLA emulation to support native Excel Pivot Tables and CUBEX functions and Power BI Tabular (DAX) models.

Data Modeling

The key to the AtScale Semantic Layer is the AtScale Semantic Model. The best way to get everyone on the same page is to have everyone speaking the same language. This ensures that there won't be conflicting answers to the same questions. A single, centralized workspace for business metrics and definitions is critical to offering one consistent, compliant view of data to business users and data scientists.

AtScale's semantic modeling tool, Design Center, works for multiple personas, including business analysts and data engineers, in a single, collaborative environment. Since AtScale Semantic Models are stored as software code in a shareable repository, models can be seamlessly integrated into your software development lifecycle (SDLC) with full CI/CD support using Git. In AtScale, every semantic object is shareable, enabling a truly decentralized but governed approach to building data products.

Benefits include:

1. Object-oriented modeling using Semantic Modeling Language (SML) promotes sharing and collaboration while drastically simplifying semantic model-building.
2. The power of a multi-dimensional engine makes even the most complex business processes easy to model.
3. Built-in Interpreters for other semantic modeling platforms like dbt, Power BI, and Looker mean that AtScale can serve any existing model regardless of the modeling language used.

Resource Orchestration

Gathering live data from multiple sources across the organization can be a long, manual process. Data engineers should create new value for the business rather than simply preparing and moving data for business reporting.

AtScale's autonomous performance optimization technology identifies query patterns and creates and manages intelligent aggregates, just like the data engineering team would. The AI-driven optimizer learns from user behavior and data relationships and takes care of data updates and changes, so business users can focus on gathering insights from data, and data engineers can concentrate on other projects. With AtScale, data access is "live" when a model is published. AtScale builds aggregates in real-time in response to user activity and automatically tunes queries without additional manual intervention.

Benefits include:

1. Automated and adaptive query performance management with AI-driven aggregate table creation and maintenance inside your data platform without the need for external resources and clusters.
2. Deep integration with your data platform that generates platform-tuned SQL to provide real-time data access without moving data.
3. Virtualized calculations using SQL and MDX to avoid unnecessary ETL for creating reporting tables and aggregations.

Governance & Metadata Management

AtScale's patented security capabilities respect native data platforms' security by supporting end-to-end user delegation and impersonation. AtScale's object-level security supports user and group access rules while providing discoverability for a 360-degree feedback experience with model designers. With integrations with enterprise data catalog and governance tools, AtScale can enforce data governance rules using AtScale's virtualized governance layer.

Benefits include:

1. Enterprise directory integration with a wide variety of IDPs for enabling single sign-on for modelers and consumers.
2. Row-level and column-level security at the modeling layer for consistent real-time policy enforcement.
3. User impersonation for supporting pass-through security to your underlying data platform.

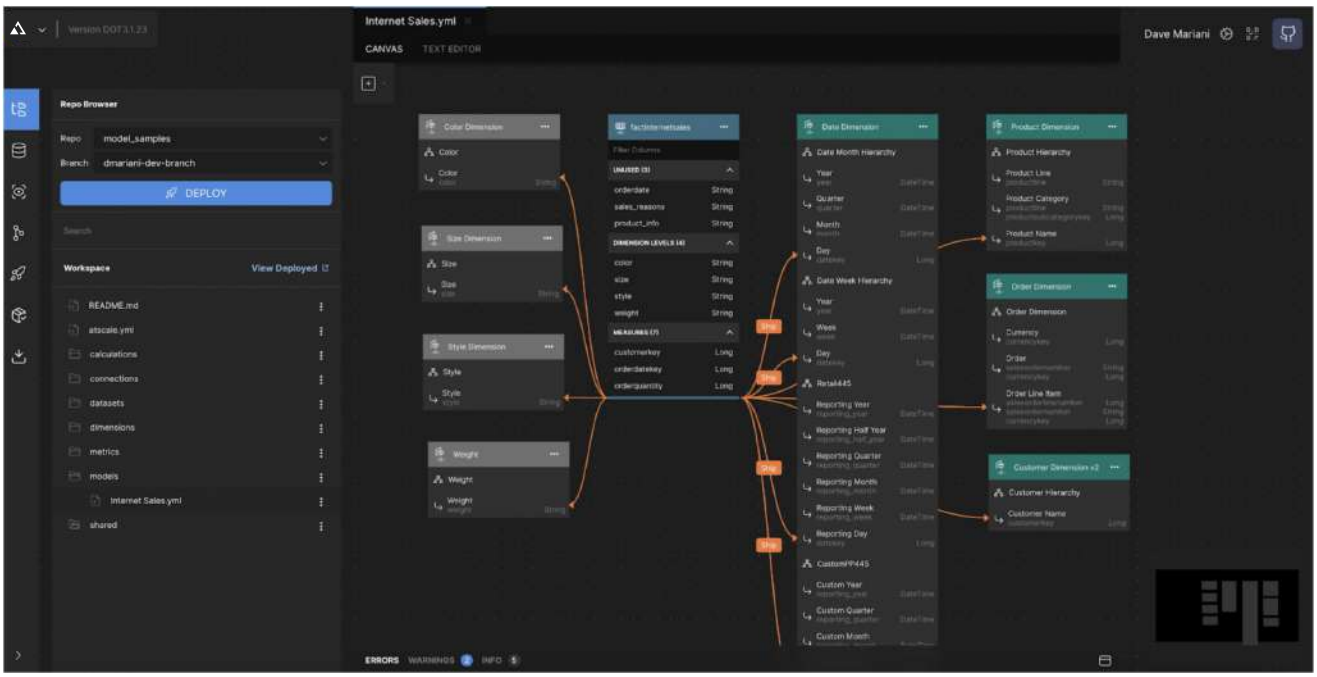
AtScale Software Component Overview

ATSCALE DESIGN CENTER

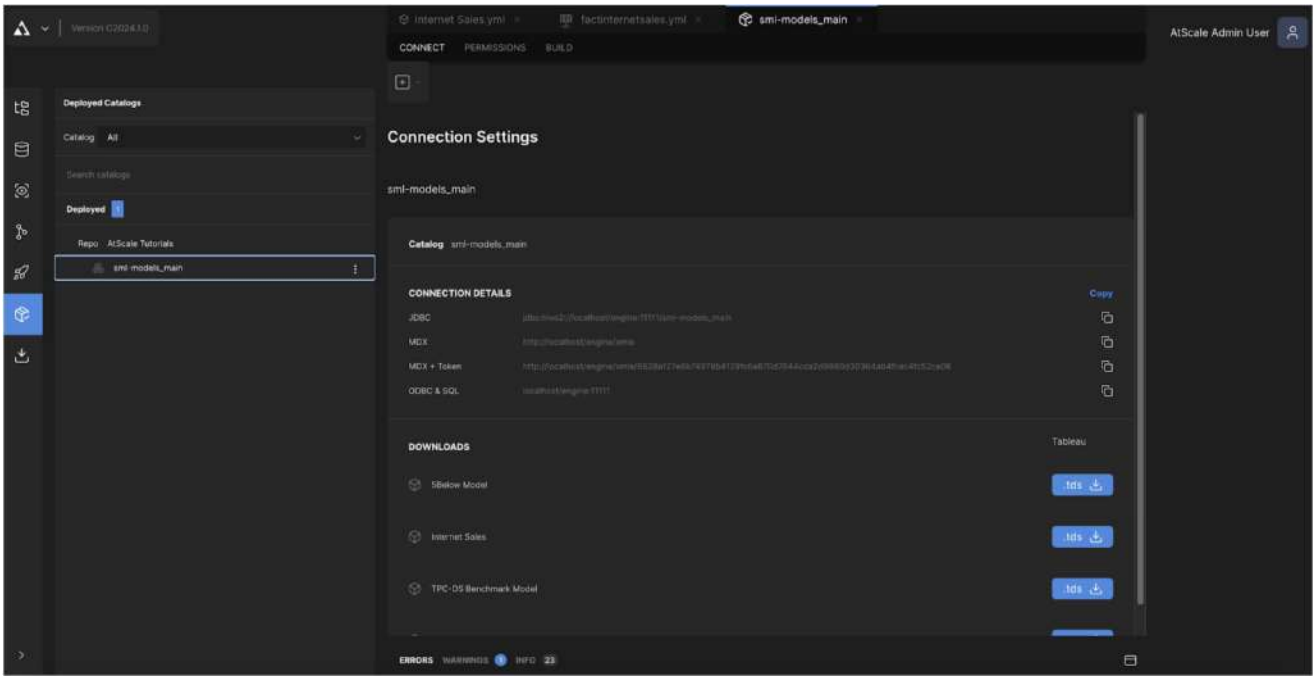
AtScale Design Center is a browser-based application that data modelers or subject matter experts (SMEs) use to create and deploy AtScale Semantic Models.

A semantic model in AtScale is a logical, business-friendly representation of a business process created with source tables through a collection of Semantic Modeling Language (SML) objects like datasets, dimensions, Measures, Hierarchies calculations, and connections.

In the following screenshots, you can see how AtScale Design Center defines models.



This image shows the main AtScale Design Center canvas. In the center, you can see blue-titled panels representing datasets (or fact tables), green-titled panels representing normal dimensions, and gray-titled panels representing degenerate dimensions (dimensions based on a fact table). The orange arrows represent relationships between the respective objects (datasets and dimensions).



Once a model is deployed in the AtScale Design Center, it is ready for consumption by BI users and data scientists. This image shows the connection instructions for Tableau, Excel, Power BI, and connection strings for AtScale's JDBC, ODBC, and MDX interfaces for custom applications.

ATSCALE QUERY ENGINE

The AtScale Query Engine is a query interface for BI, AI/ML tools, and custom applications. Tools can connect to AtScale via one of the following protocols:

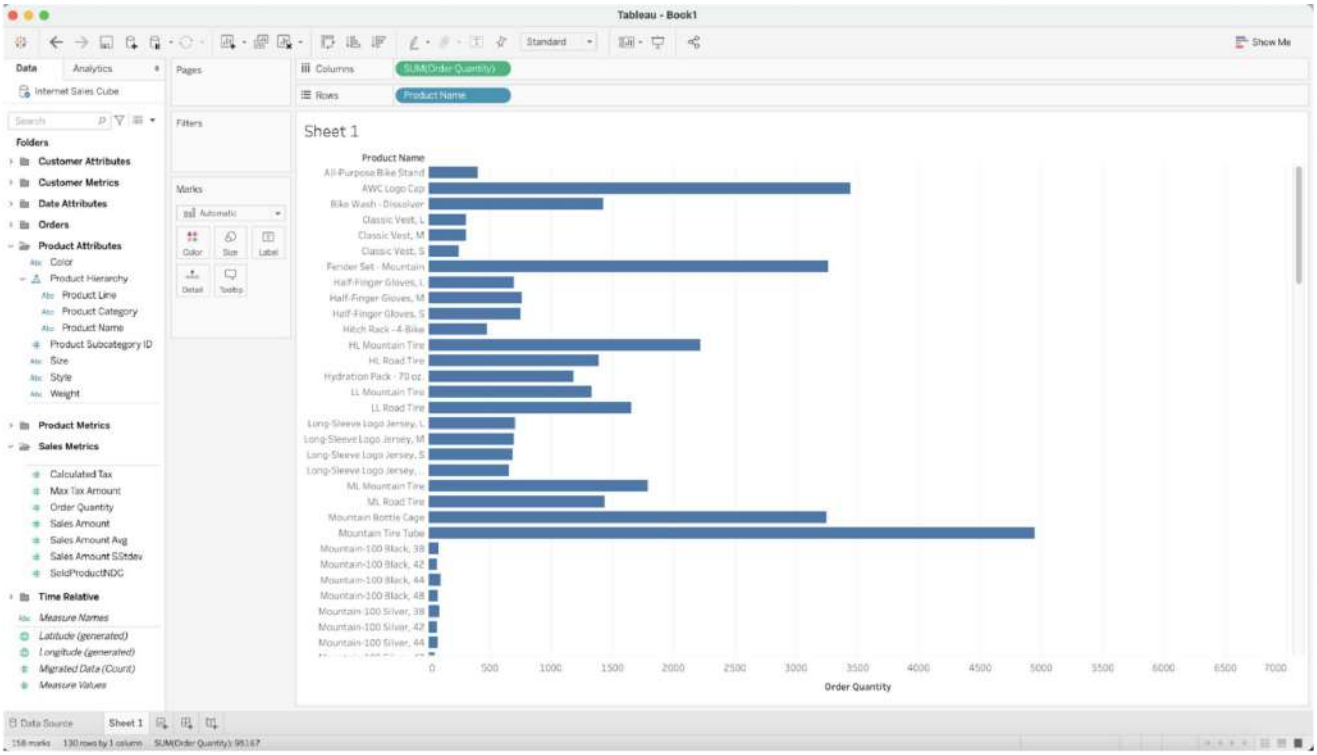
1. ODBC/JDBC (SQL)
2. XMLA (MDX or DAX)
3. Python
4. REST

The AtScale engine appears as a PostgreSQL table for the tools that speak SQL. AtScale appears as a SQL Server Analysis Services (SSAS) cube for the tools that talk to MDX or DAX. For applications speaking REST or Python, AtScale appears as a web service.

AtScale's Semantic Layer provides the same logical view of business-friendly data regardless of the BI and AI/ML tools. Users can interact with data using the same dimensions, hierarchies, and measures defined in the Design Center. AtScale delivers data as a service to all data consumers without degrading the user experience for their respective consumption tools.

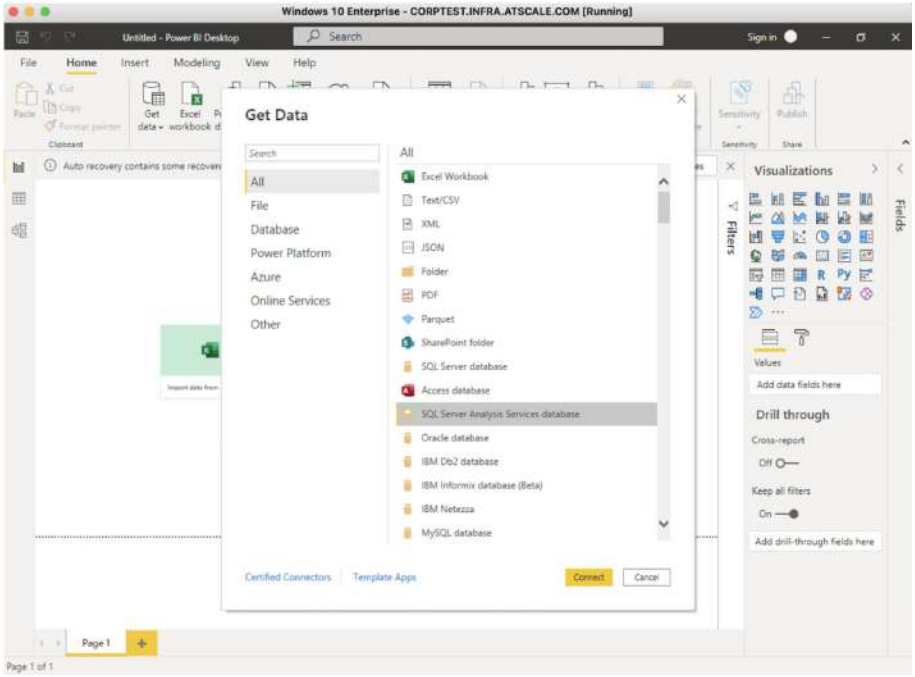
The following screenshots show how the AtScale Query Engine appears in various consumption tools.

TABLEAU

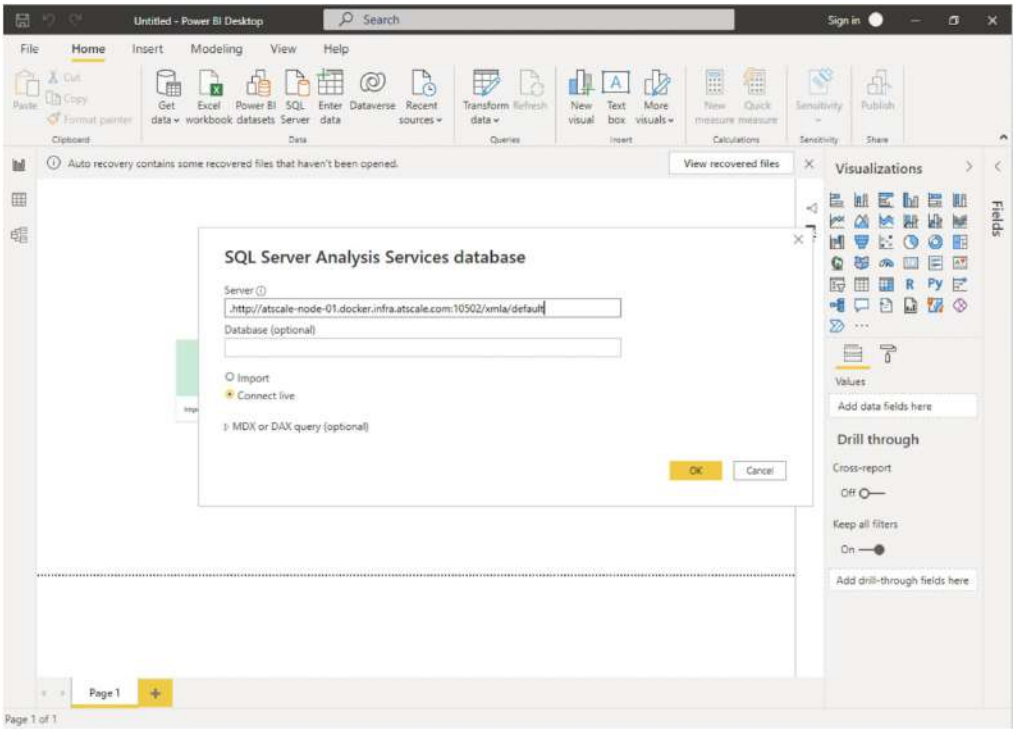


This image shows how the AtScale Semantic Layer for the “Internet Sales Cube” appears to consumers in Tableau. It also shows the results of the “Order Quantity by Product Name” query.

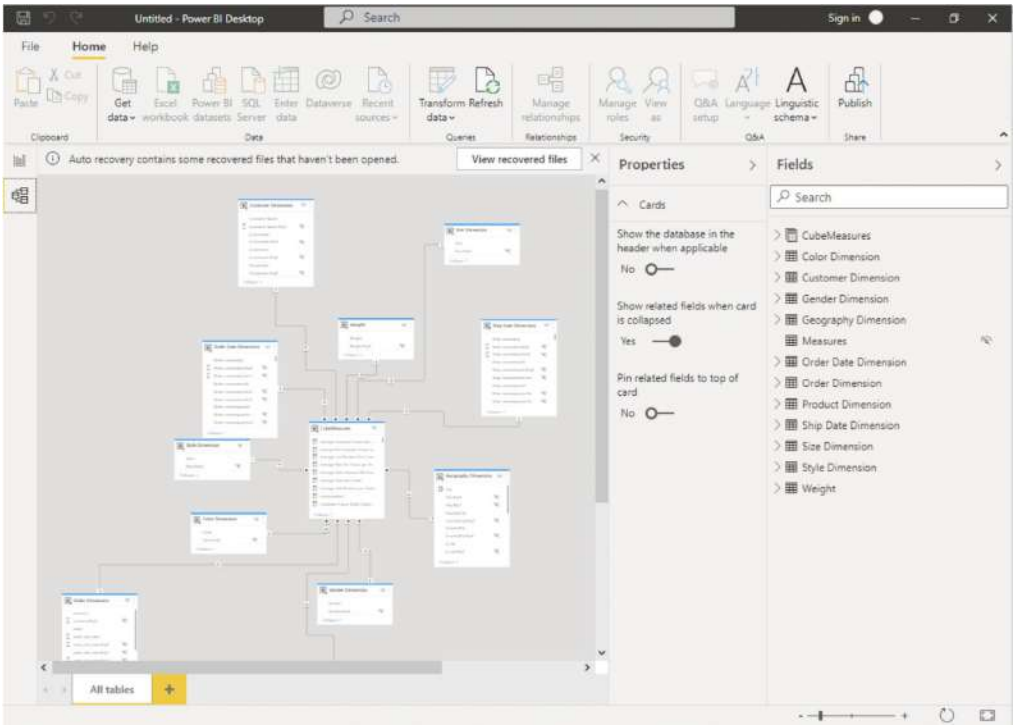
POWER BI



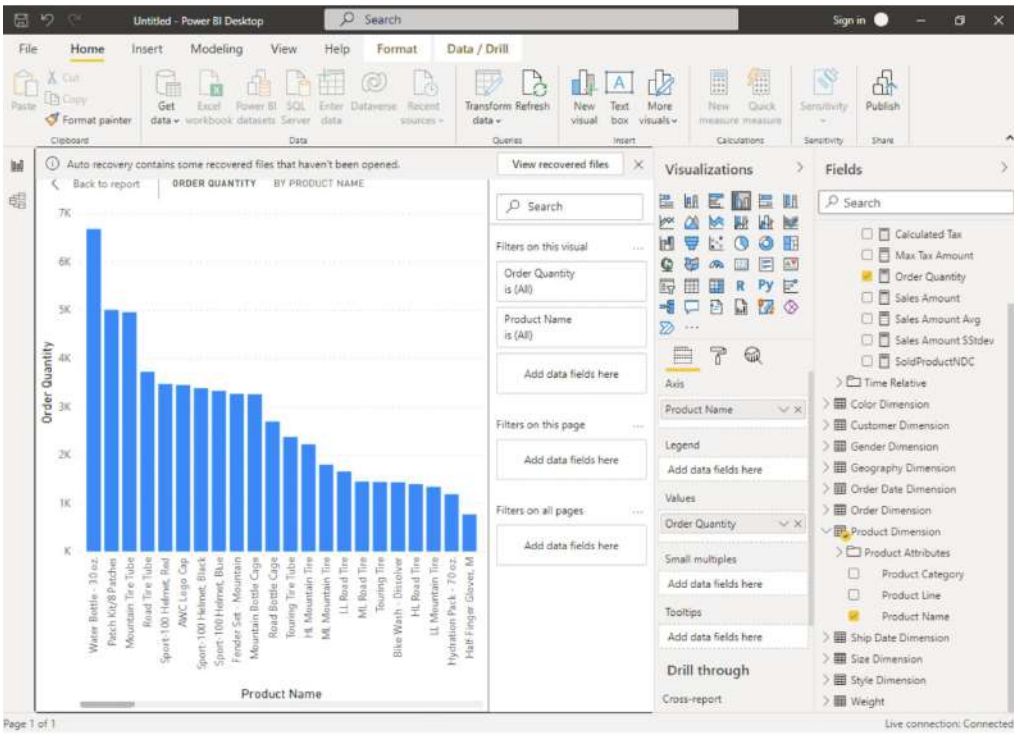
This image shows how Power BI users can leverage the built-in SQL Server Analysis Services (SSAS) connector. This means there is no need to install custom client-side drivers to access AtScale models.



This image shows how Power BI connects in “Live” query mode to the AtScale Query Engine.

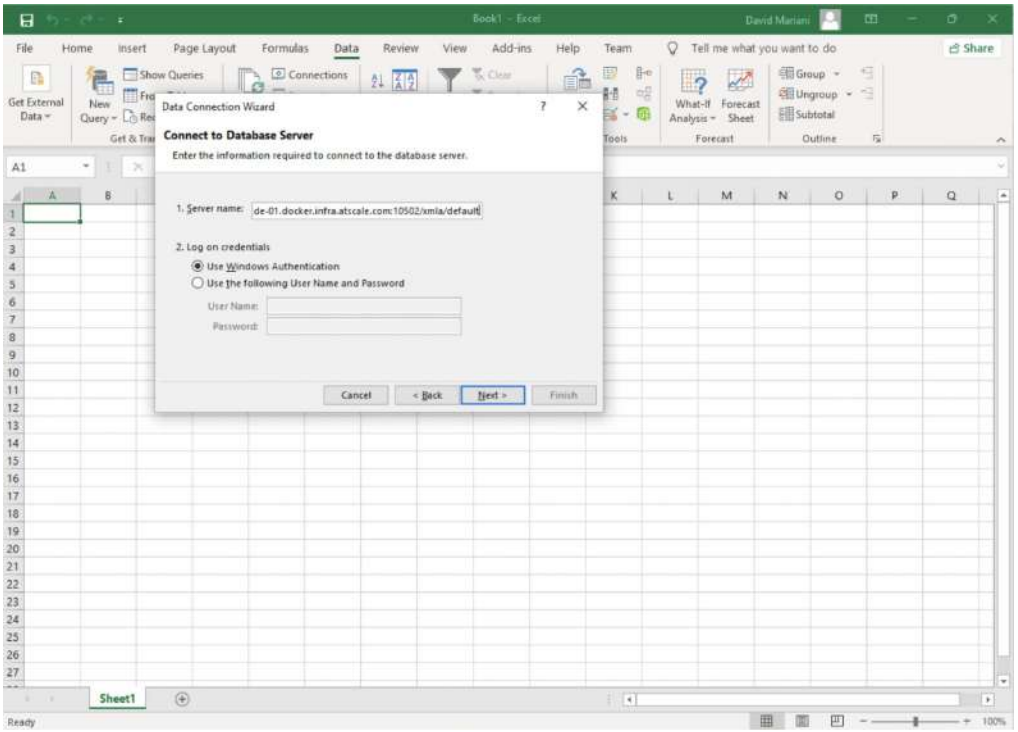


In this image, you can see how the AtScale semantic layer for the “Internet Sales Cube” is inherited automatically in Power BI, requiring no additional models for analytics consumers.

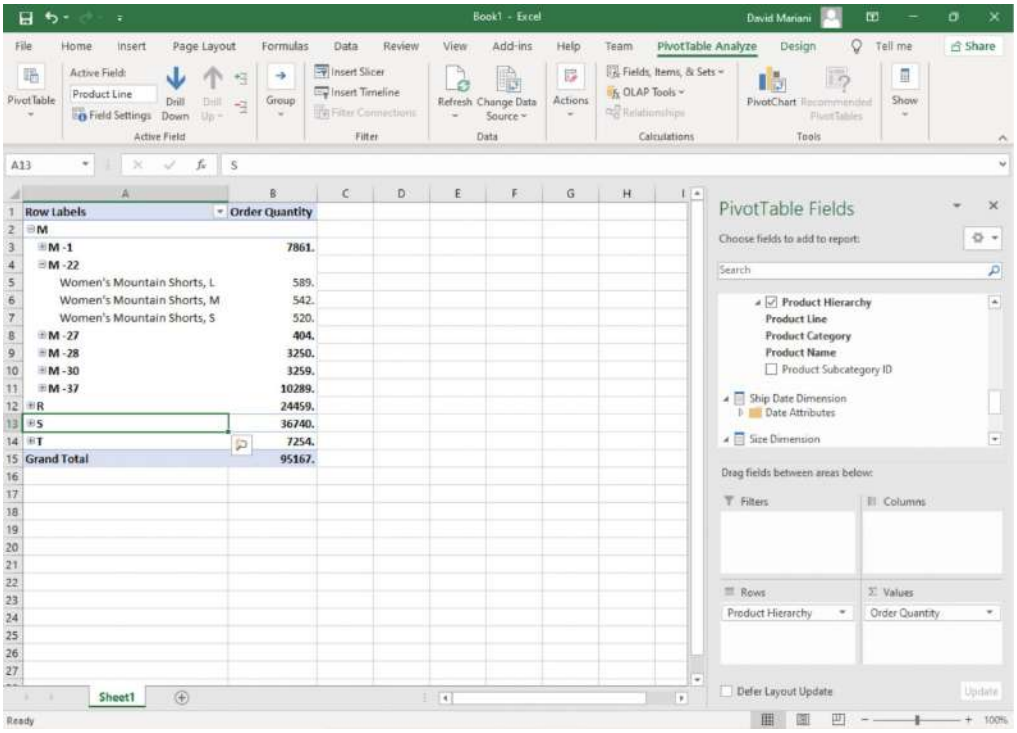


This image shows how the AtScale semantic layer for the “Internet Sales Cube” appears to consumers in Power BI. This image shows the results of the “Order Quantity by Product Name” query.

EXCEL

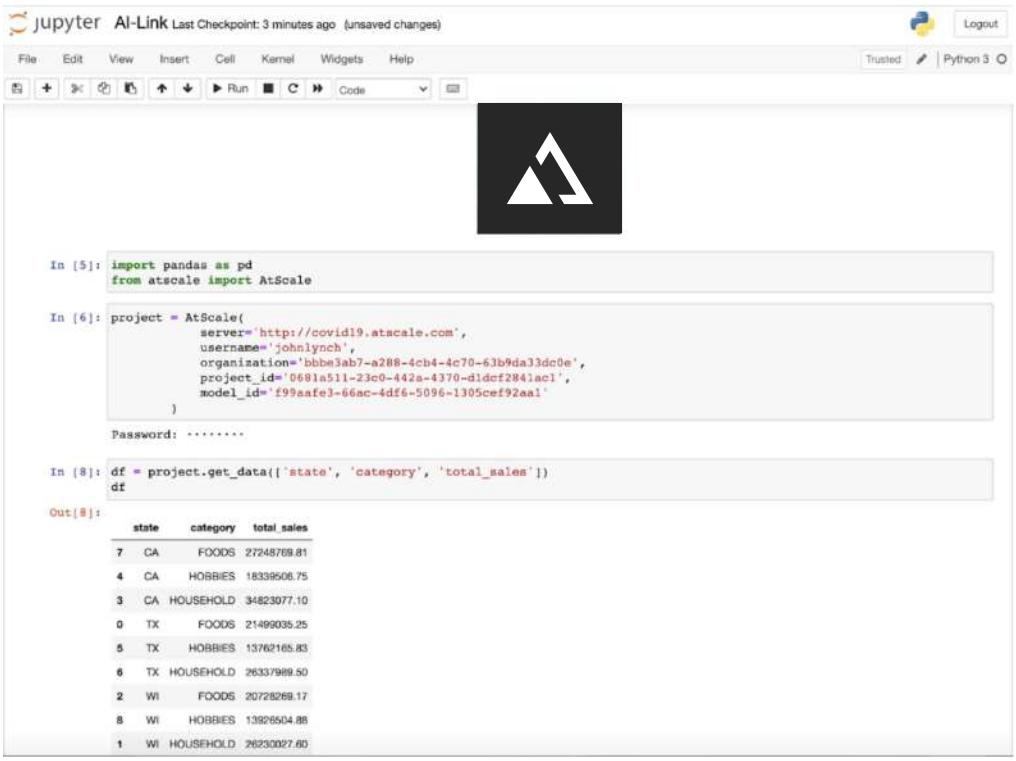


This image shows how Excel users can leverage the built-in SQL Server Analysis Services (SSAS) connector. This means there is no need to install custom client-side drivers to access AtScale models. Also, we are connecting using Window Authentication (Active Directory).



This image shows how the AtScale semantic layer for the “Internet Sales Cube” appears to consumers in Excel. It also shows the results of the “Order Quantity by Product Name” query.

JUPYTER



This image shows how the AtScale semantic layer appears to a data scientist using a notebook and AtScale’s Python interface.

Deployment

AtScale installs on Kubernetes via Helm chart. Once installed, Kubernetes provides the cluster on which the AtScale services run and automates management, scaling, and failover for the services.

The AtScale Developer Community Edition is installed on Docker, which provides the platform for running and managing the AtScale container.

The AtScale instance serves as a query endpoint for BI/AI tools and a modeling endpoint for AtScale Design Center, a browser-based design environment for creating and managing models.

Supported Business Intelligence Tools

Excel and Power BI contain the required drivers. SQL tools such as Tableau require an additional Hive or PostgreSQL driver.

TOOL	VERSION(S)	CONNECTION TYPE
Tableau Desktop and Server	2023.3	Thrift SQL
Looker	24.2	Thrift SQL
Excel	2021, 2019, 365	XMLA (MDX)
Power BI	February 2024	XMLA (DAX)
Power BI Service	N/A	XMLA (DAX)

Unsupported Tools: The following BI tools have basic connection and query support, however, they are not fully supported: Microstrategy, Business Objects, Cognos, Saiku, and Spotfire.

User Management and Authentication

The Identity Broker service manages AtScale users. For production environments, it can be paired with your organization's identity provider or LDAP server. AtScale supports the following identity providers and LDAP servers:

- Okta (with OpenID Connect or SAML 2.0)
- Microsoft Entra ID (with OpenID Connect or SAML 2.0)
- Windows Active Directory (supports Kerberos authentication protocol)

Container Platforms

AtScale:

- Kubernetes

AtScale Developer Community Edition:

- Docker

Frequently Asked Questions

What do I need to deploy AtScale?

AtScale can be deployed with Docker Compose or Kubernetes, supported by a Helm chart. You need to configure AtScale to point to a supported data platform as listed in the **Integrations** section of this document. While not required, you will also want to configure AtScale to access your directory service (AD/LDAP) and your external load balancer for High Availability (HA) configurations. AtScale installation requires at least one Linux server or virtual machine, and some basic prerequisites are required to install the AtScale software. You may need the appropriate JDBC/ODBC drivers for client tool access if they aren't already installed. No additional driver is necessary for Excel, Power BI, or tools that use the XMLA (MDX, DAX) protocol.

Is there a trial and/or open-source version of AtScale?

AtScale now offers a Developer Community Edition of AtScale that is free to download. The Developer Community Edition is a fully featured, free version of AtScale's industry-leading semantic layer platform, promoting collaboration and versatility. As users build and deploy more models, scaling is simplified, paving the way for users to transition to commercial AtScale for the enterprise seamlessly.

Get access here.

How does AtScale interact with my data platform?

AtScale is a client to your data platform(s) and will generate optimized, platform-specific SQL based on the AtScale model defined in the AtScale Design Center.

Once a model is deployed, it is immediately available for BI and/or AI/ML activity. No pre-processing or data movement is required when publishing a model. Data consumers can connect to the AtScale engine via ODBC/JDBC (SQL), XMLA (MDX, DAX), REST, or Python interfaces and begin querying the model.

AtScale intercepts inbound queries from the end user's query tools and rewrites them for execution on a data platform, leveraging any available AtScale-managed aggregates that would benefit the user's query.

Simultaneously, AtScale's machine learning algorithms monitor user activity and manage aggregations to automatically optimize query performance. AtScale creates, manages, and stores aggregate tables in a schema in the underlying data platform(s).

What are the options for aggregate creation?

Aggregates may be triggered in three ways:

1. **Demand-based Aggregates** are generated heuristically based on user query behavior.
2. **Predictive Aggregates** are generated proactively based on model design. For example, dimensional aggregates may be generated to facilitate fast lookups for building reports.

3. **User-defined Aggregates** are defined by the AtScale Design Center modeler and stored inside the AtScale model. Users can specify combinations of dimensions and measures to design an aggregate manually, and these aggregates will automatically be built when the model is published.

In addition to these types, settings are available for adjusting behavior and thresholds to create demand and prediction-based aggregates.

How are the acceleration structures managed and kept current?

There are three methods of controlling how and when the acceleration structures are refreshed.

1. Aggregates may be refreshed on a time or calendar basis using AtScale's built-in scheduler.
2. Aggregates may be refreshed using AtScale's file watcher utility on a file-trigger basis. This method is often used in conjunction with an ETL pipeline to trigger a refresh upon the completion of an ETL flow.
3. Aggregates may be refreshed using AtScale's REST API. As with the file trigger option, this method is often used in conjunction with an ETL pipeline workflow.

Aggregates can be updated either incrementally or in full refresh mode. Incremental updates allow for appending new or changed data, whereas a full refresh rebuilds the aggregates from scratch.

About AtScale

AtScale's semantic layer platform accelerates data-driven insights, simplifying and extending BI and data science capabilities. Their platform empowers enterprise customers to democratize data, enabling self-service BI and agile analytics for impactful decision-making. For more information, please visit www.atscale.com and follow us on LinkedIn.