ATSCALE

Technical Overview

ATSCALE

TABLE OF CONTENTS

ATSCALE

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 ands 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 take advantage of 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 the security risks associated with them.

TURBOCHARGE ANALYTICS AND MACHINE LEARNING INITIATIVES
Enterprises can Instantly integrate new data sources without manual data 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.
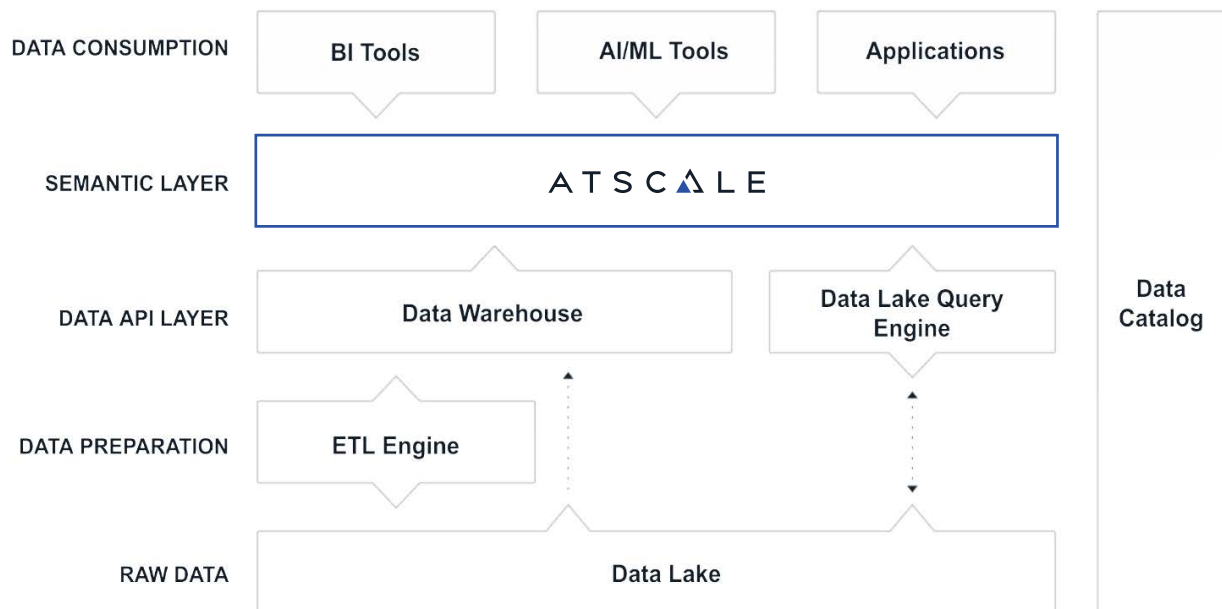
CONDUCT INTERACTIVE AND MULTIDIMENSIONAL ANALYSES
Using business users' preferred BI tools, whether that is Excel, Power BI, Tableau, or something else.

GET CONSISTENT ANSWERS ACROSS DEPARTMENTS AND BUSINESS UNITS
Via AtScale's Semantic Layer that standardizes queries regardless of BI tool or query language.
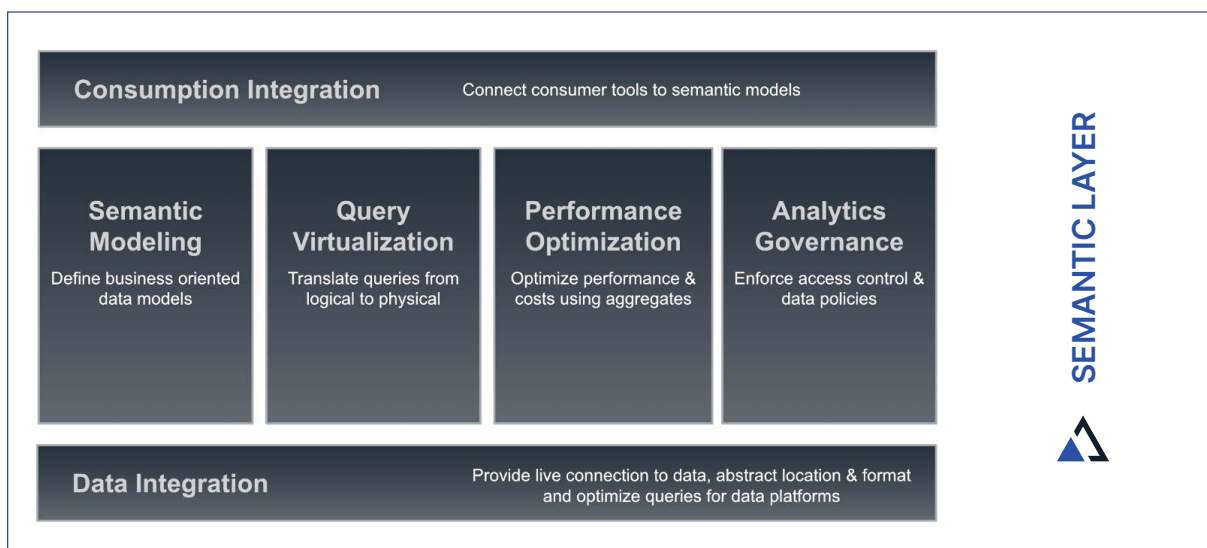
Where Does AtScale Fit in the Analytics Stack?

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.

DATA CONSUMPTION	BI Tools	AI/ML Tools	Applications	
SEMANTIC LAYER	ATSCALE			Data Catalog
DATA API LAYER	Data Warehouse		Data Lake Query Engine	
DATA PREPARATION	ETL Engine			
RAW DATA	Data Lake			

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.

In the following sections, we will describe the inner workings and benefits of each component.

ATSCALE

Consumption Integration

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

Benefits include:

Multimodal

- Supports tabular and multi-dimensional clients
- Supports BI and AI/ML workloads
- Application and integration friendly via REST, JDBC, ODBC, MDX and DAX

Lightweight Footprint

- No custom client-side drivers needed
- Wire compatible with Microsoft SSAS protocol (Excel, Power BI)
- Live connection to data without extracts or data imports

Client Optimized

- Optimized SQL interface for Tableau
- Native DAX support for Power BI and live pivot tables for Excel
- LookML integration for Looker

Semantic Modeling

The key to the AtScale semantic layer is the AtScale 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 key to offering one consistent, compliant view of data to both business users and data scientists alike.

AtScale's semantic model unifies semantic definitions and metrics for data and makes it available in one location for BI, AI/ML and applications. It works on data anywhere whether it's in a data lake or a data warehouse.

Benefits include:

Object Oriented
- Reuseable cubelets for conformance
- Library to promote sharing across models
- Multi-pass calculation for complex expressions

Multi-source
- Blend data from multiple sources in a model
- Smart push down for optimal performance
- Data platform optimized

Open
- XML-based markup language
- Programmable APIs for model creation
- Native SQL for full platform functionality

Query Virtualization

AtScale's data virtualization automates the sourcing, curation and modeling of data on premises or in the cloud. It blends live data from multiple data sources into virtual, logical views. Virtualization makes IT more agile with the ability to store data in the most suitable platforms while providing the flexibility to adopt new platforms in the future without re-architecting their stack or disrupting their downstream data consumers.

AtScale's Data virtualization provides access to enterprise data by functioning as an abstraction layer on top of a variety of data platforms but without manually moving data.

Benefits include:

Built for BI
- Supports tabular and multi-dimensional modeling
- Supports SSAS compatible MDX and DAX (Excel)
- Time Intelligence, Hierarchies, Cell-level calcs

Graph-based Planner
- Node-based query planning and optimization
- Dialect and data platform optimized
- Scalable to thousands of dimensions

Integration First
- Query delegation and directory support
- Integration with enterprise data catalogs
- Dimensional sharing model (Conformed Dims)

Performance Optimization

Gathering live data from multiple sources across the organization can be a long, manual process. Data engineers should be creating 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 do. 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 focus on other projects. With AtScale, the moment a model is published, data access is "live". AtScale builds aggregates in real-time in response to user activity and automatically tunes queries without additional manual intervention.

Benefits include:

Machine Learning Powered
- Uses query behavior patterns
- Leverages HLL++ to computer statistics
- Proactively optimizes at model time

Adaptive, Always Working
- Automatically prunes, splits, merges aggregates
- Automatically adapts to data and model changes
- Supports full and incremental updates

Platform Optimized
- Optimizes locale via preferred agg storage (PAS)
- Platform specific partitioning and ordering
- HLL++ libraries adapted for each platform

Analytics Governance

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:

Environmentally Smart
- Enterprise Directory Integration (AD/Octa/OAuth)
- Secure Cluster Support (Kerberos, Ranger, files)
- Secure Data Transport Protocols (TLS)

Just in (query) Time
- Configurable delegation query execution
- Impersonation throughout (including client)
- Query time row level masking for user/groups

Object Oriented
- Role-based, object level control for users/groups
- Column masking in query tools (Perpectives)
- Multi-tenancy support via organization level

Data Integration

First generation data virtualization was not designed for the large analytical workloads that are typical of today's BI and AI use cases. AtScale's deep expertise in multidimensional analytics along with federated query processing provides unparalleled support for BI and AI tools alike. AtScale speaks to data lakes and data warehouses with data platform optimized SQL so performance is as fast or faster than hand-written queries. Rather than processing data locally, the AtScale engine pushes queries down to the underlying data platform to eliminate data movement and scale performance along with the data platform, without the need of managing a separate analytics infrastructure.

Benefits include:

Environmentally Smart
- Enterprise Directory Integration (AD/Octa/OAuth)
- Secure Cluster Support (Kerberos, Ranger, files)
- Secure Data Transport Protocols (TLS)

Just in (query) Time
- Configurable delegation query execution
- Impersonation throughout (including client)
- Query time row level masking for user/groups

Object Oriented
- Role-based, object level control for users/groups
- Column masking in query tools (Perpectives)
- Multi-tenancy support via organization level

AtScale Software Component Overview

AtScale Design Center

AtScale Design Center is a browser-based application that subject matter experts (SMEs) use to create and publish AtScale models. AtScale Design Center is organized around the following object model:

ORGANIZATIONS: The 'top-level' structure contains sets of users, groups, roles, permissions, and more. AtScale instances will have at least one Organization. Additional Organizations can be created and they are mutually exclusive. Nothing is shared between Organizations.

PROJECTS: A Project is a collection of one or more AtScale models. Each project contains a shared Library with objects like Datasets and Dimensions that can be shared with other models in that project. There is no limit to how many Projects you can create or how many objects can be shared.

MODELS: A Model is created within a Project and can use shared objects from that Project's Library. A Model is a collection of Datasets, Dimensions, Measures, Hierarchies, and Calculations along with their Relationships that form the basis of a virtual, multidimensional view of your source data.

Models in AtScale visually look like a 'star' or 'snowflake' schema. However, there is no requirement for any particular physical data structure or layout: data can be normalized or denormalized or a little bit of both.

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

The above screenshot shows the main AtScale Design Center canvas. In the center, you can see blue titled panels that represent datasets (or fact tables), green titled panels that represent normal dimensions and gray titled panels that represent degenerate dimensions (dimensions based on a fact table). The orange arrows represent relationships between the respective objects (datasets and dimensions). In the right hand panel, you can see a preview of the dimensions, measures and hierarchies (the logical model) that the end query consumer will see.

In the above screenshot, you can see the AtScale Design Center data wrangling view for the "factinternetsales" Snowflake table. You can apply transformations and create new calculations for the semantic model.

This image shows a "Customer" dimension built from a base table called "dimcustomer" and connected to two other dimensions, "Geography" and Gender". By building conformed dimensions, new models are both easy to create and consistency is insured across models.

Once a model is published in AtScale Design Center, it is ready for consumption by BI users and data scientists. In this image, you can view the connection instructions for Tableau, Excel, Power BI and connection strings for AtScale's JDBC, ODBC and MDX interfaces for custom applications.
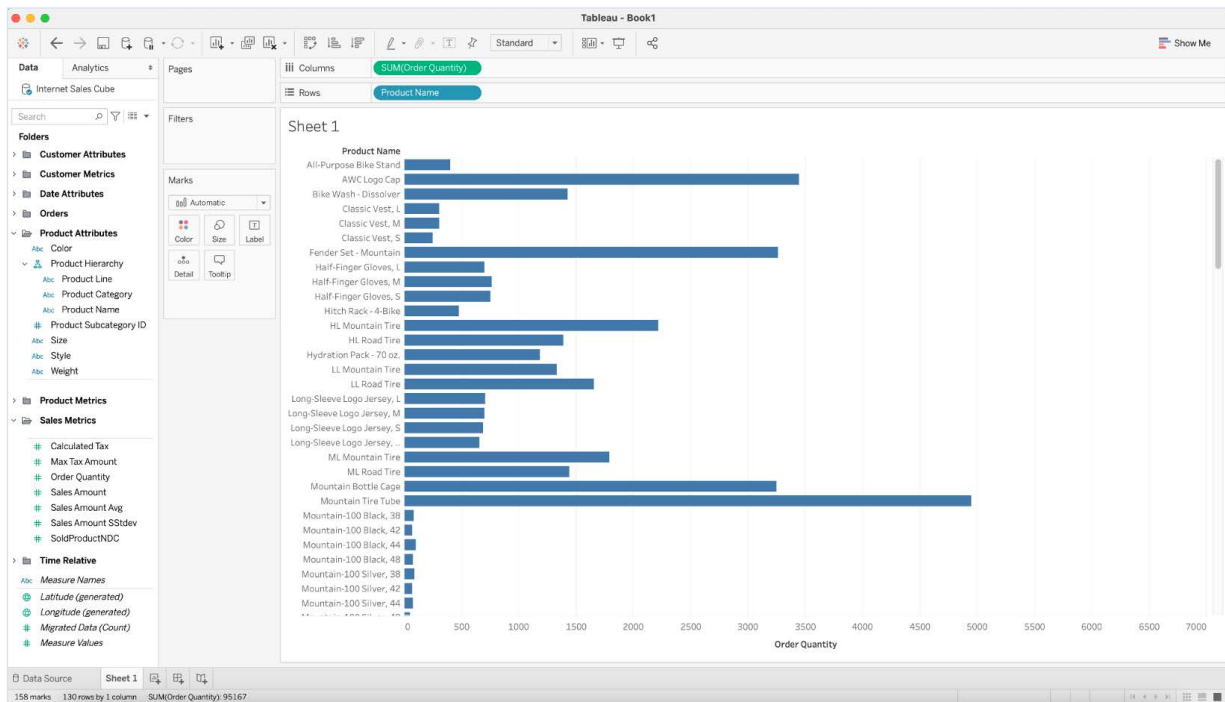
AtScale Query Engine

The AtScale Query Engine acts as a query interface for business intelligence, AI/ML tools and applications. Tools can connect to AtScale via one of the following protocols:

1. ODBC/JDBC (SQL)

2. XMLA (MDX or DAX)
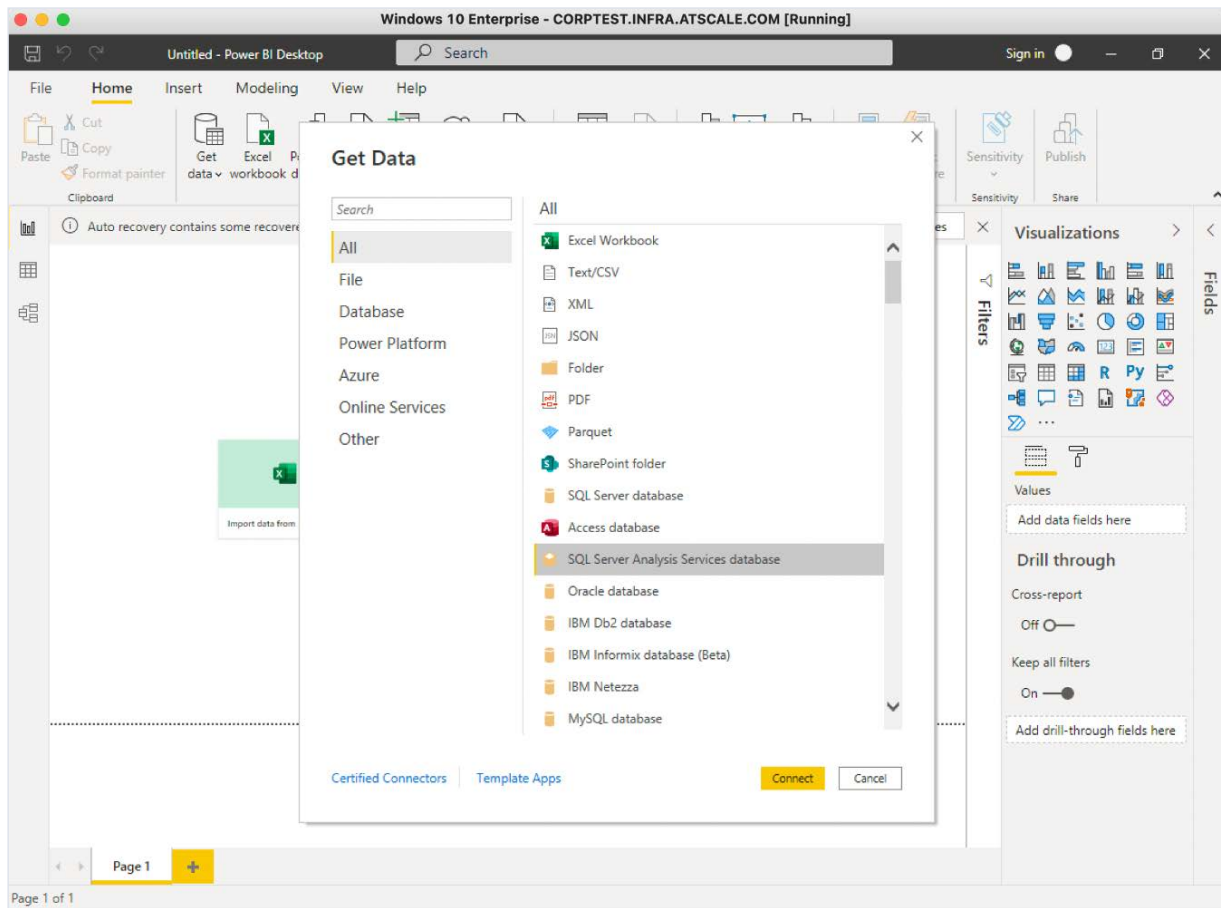
3. PYthon

4. REST

For the tools that speak SQL, the AtScale engine appears as a Hive SQL warehouse. For the tools that speak MDX or DAX, AtScale appears as a SQL Server Analysis Services (SSAS) cube. For applications speaking REST or Python, AtScale appears as a web service. The following screenshots show how the AtScale Query Engine appears in a variety of consumption tools.

AtScale's semantic layer provides the same logical view of business-friendly data regardless of the BI and/or AI/ML tool. Users can interact with data using the same dimensions, hierarchies, and measures defined in the Design Center. With AtScale, data is delivered as a service to all data consumers without any restrictions to share and collaborate.
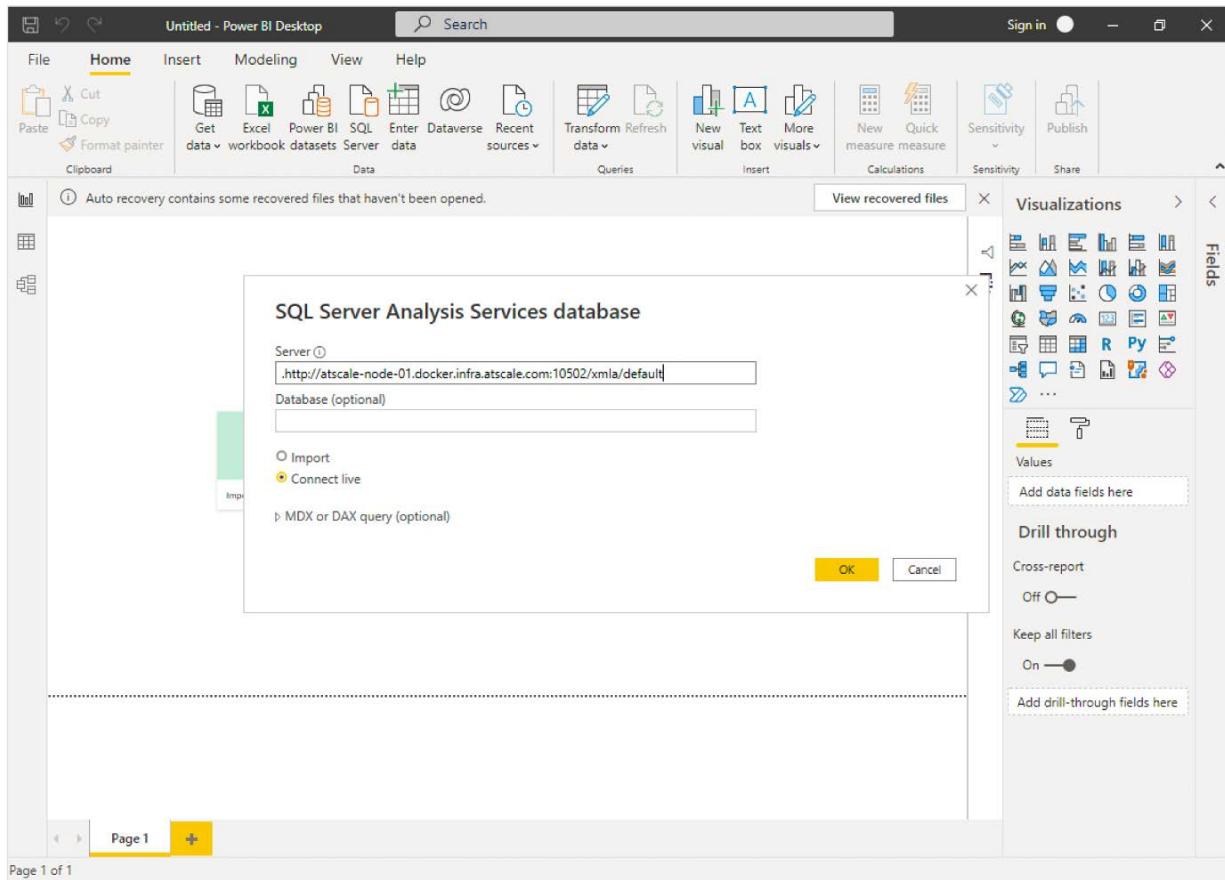
In this image, you can see how the AtScale semantic layer for the "Internet Sales Cube" appears to consumers in Tableau. This image shows the results of "Order Quantity by Product Name" query.
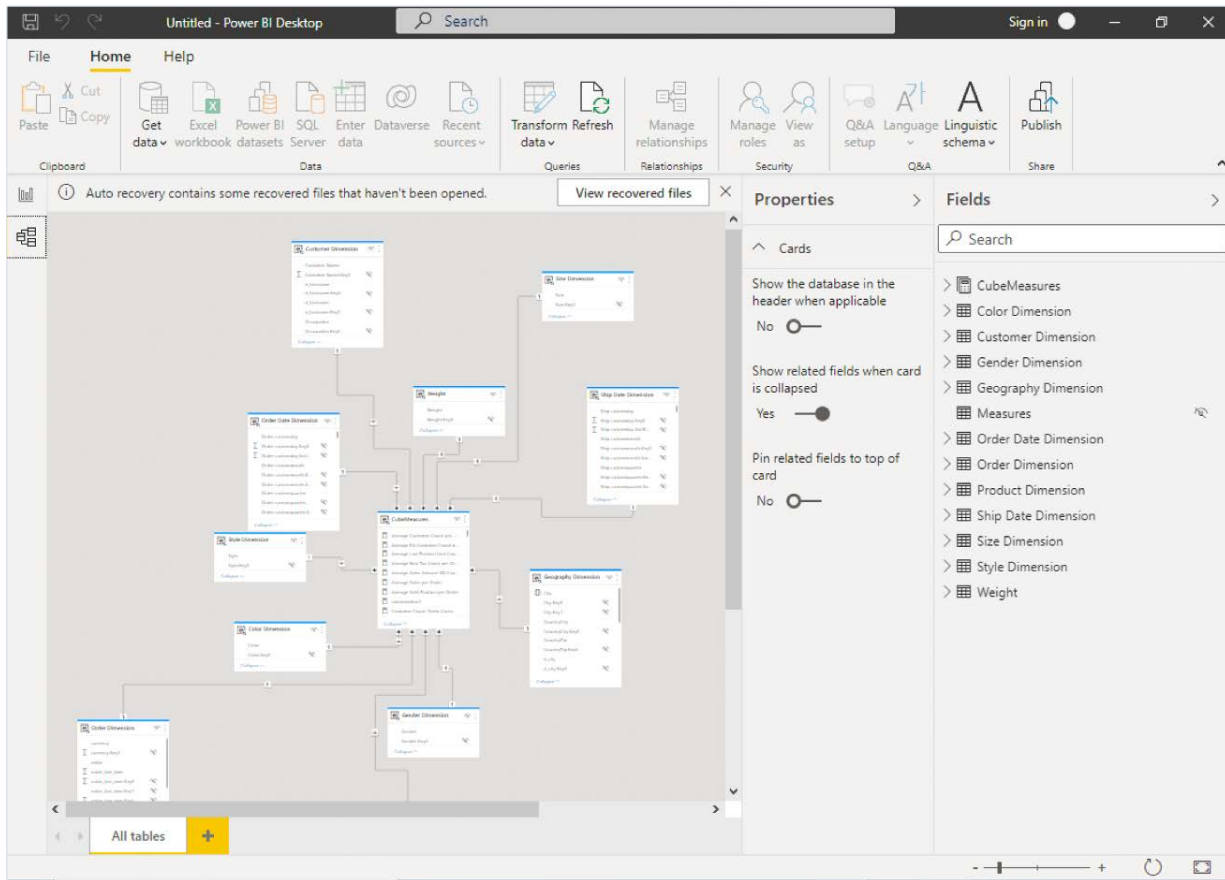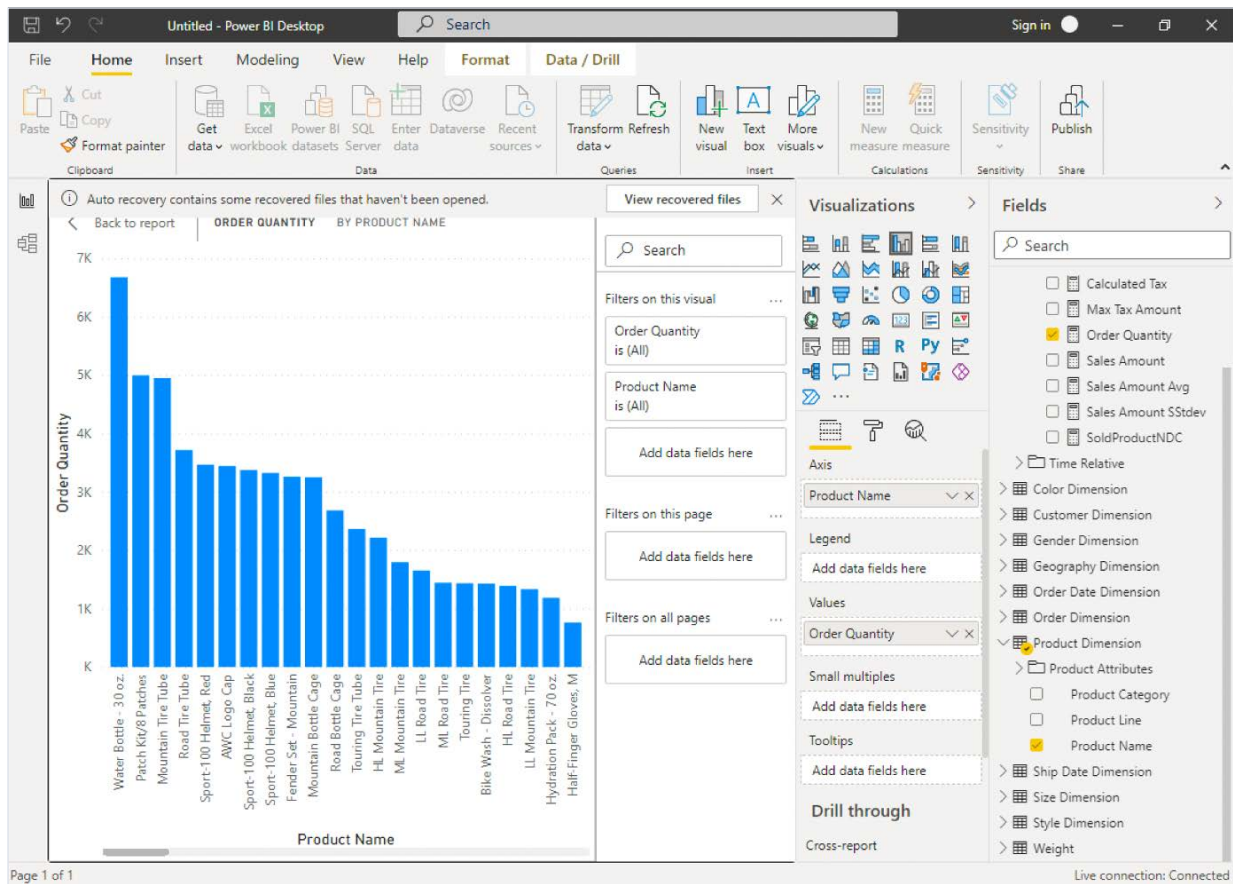
In this image, you can see how the Power BI users can leverage the built-in SQL Server Analysis Services (SSAS) connector. This means that there is no need to install custom client-side drivers to access AtScale models.
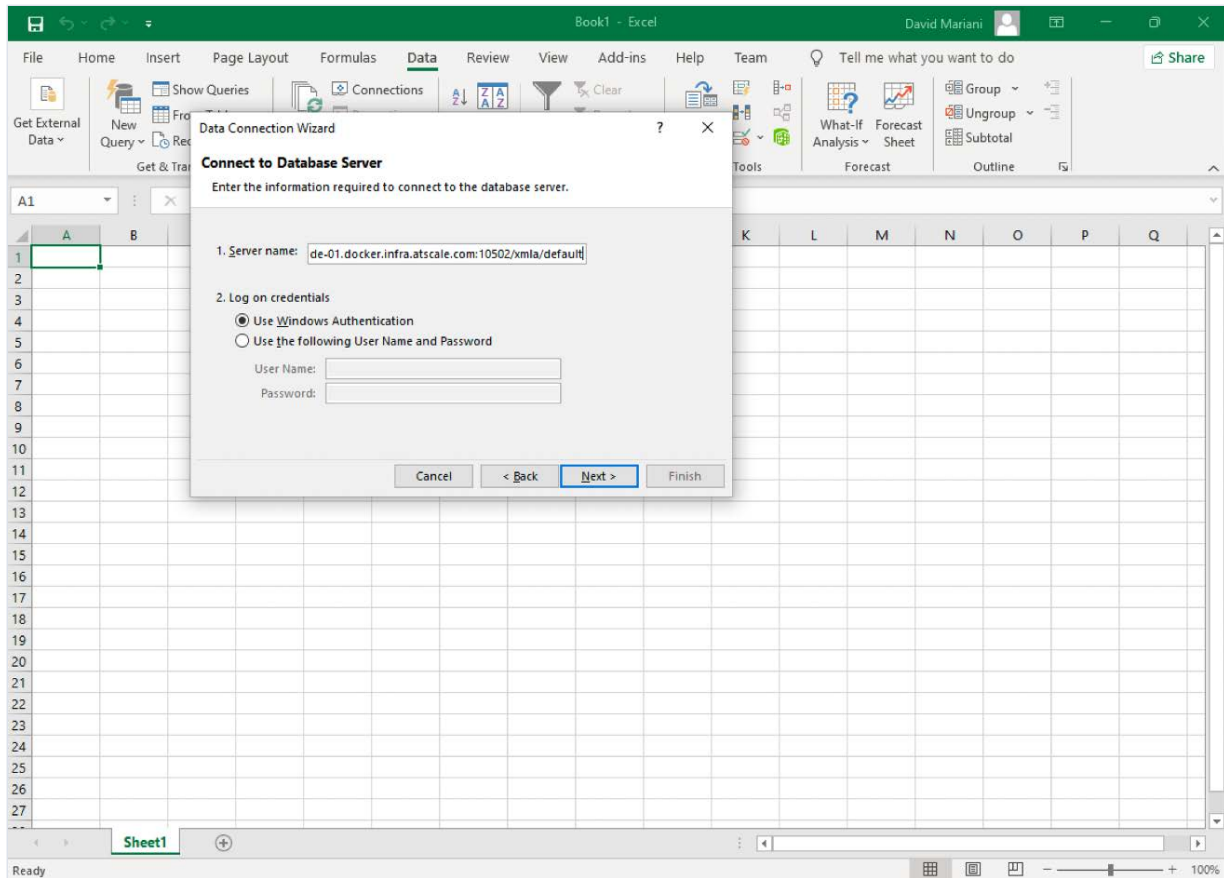
In this image, you can see how Power BI connects in "Live" query mode to the AtScale Query Engine.

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.

In this image, you can see how the AtScale semantic layer for the "Internet Sales Cube" appears to consumers in Power BI. This image shows the results of "Order Quantity by Product Name" query.
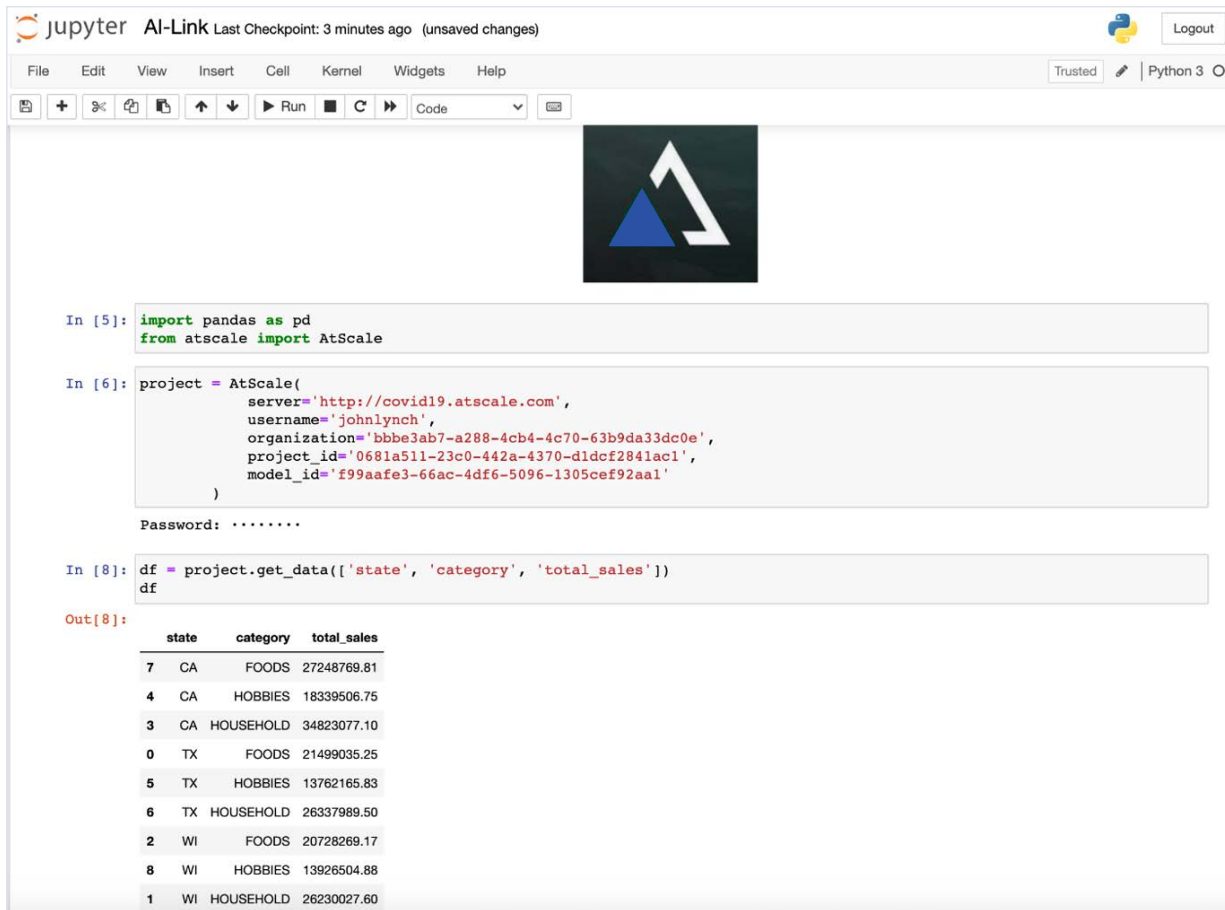
In this image, you can see how the Excel users can leverage the built-in SQL Server Analysis Services (SSAS) connector. This means that there is no need to install custom client-side drivers to access AtScale models. Also, notice that we are connecting using Window Authentication (Active Directory).

In this image, you can see how the AtScale semantic layer for the "Internet Sales Cube" appears to consumers in Excel. This image shows the results of "Order Quantity by Product Name" query.

In this image, you can see how the AtScale semantic layer appears to a data scientist
using a notebook and AtScale's Python interface.

Deployment

AtScale micro-services install on a Linux server or virtual machine within any environment, either on-premise or in the cloud using a RPM. The AtScale instance serves both as a query endpoint for BI/AI tools and modeling endpoint for AtScale Design Center, a browser-based design environment for creating and managing models.

For the query federation option, AtScale also installs and manages a Spark service to manage heterogeneous joins across multiple data sources.

This image shows a graphical representation of the AtScale service components.

This image shows the AtScale Design Center Orchestrator console for managing AtScale components and services.

The AtScale platform consists of multiple services:

1. **Agent**
   Installed on a virtualization fabric node to communicate with the virtualization service.

2. **Balancer**
   Internal load balancers for routing traffic for High Availability (HA).

3. **Coordinator**
   Installed at least 3 nodes for managing High Availability (HA).

4. **Database**
   AtScale internal database for storing application and configuration data.

5. **Directory**
   Internal LDAP directory used if an external directory is not defined.

6. **Engine**
   AtScale query engine service.

7. **Health**
   Health check service.

8. **Ingress**
   Bridge service for virtualization configuration services.

9. **Modeler**
   AtScale Designer Center service.

10. **Servicecontrol**
    AtScale services manager.

11. **Virtualization_listener**
    Virtualization fabric listener service.

12. **Virtualization_supervisor**
    Virtualization fabric supervisor service.

13. **Virtualization_worker**
    Virtualization fabric worker.

Supported Business Intelligence Tools

Excel and Power BI contain the required drivers. SQL Tools such as Tableau must use one of the supported Hive drivers listed below. AtScale requires a Hive driver because it emulates the Hive SQL interface.

Tool	Version(s)	Connection Type
Tableau Desktop and Server	2021.1.0, 2020.4.5, 2020.3.1, 2020.2.2, 2020.1.5, 2019.4.0	Thrift SQL
Looker	21	$32.00
Excel	2016, 2013, 2010	XMLA (MDA)
Power BI	NA	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, Spotfire.

Experimental Tools: AtScale has experimental support for Qlik. To use Qlik with AtScale, you must have a special AtScale license. Ask your sales representative about obtaining a Qlik-enabled license.

Directory Services

- Windows Active Directory 2012 (Supports Kerberos authentication protocol)

- Google Directory Service (NOTE: Does not support Kerberos, which is required for Tableau Single Sign-on)

- LDAP 2 compatible directory service providers

Operating System Support

- CentOS 6

- CentOS 7

- Ubuntu 16.04

- Ubuntu 18.04

- Ubuntu 20.04

Frequently Asked Questions

What do I need to deploy AtScale?

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. For AtScale installation, at least one Linux server or virtual machine is required with some basic prerequisites to install the AtScale software. For client tool access, you may need the appropriate JDBC/ODBC drivers 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?

You need to configure AtScale to point to a supported data platform as listed in the AtScale supports a proof-of-concept trial. Please contact us to discuss your use case and/or project needs to determine if a proof-of-concept trial would be appropriate.

How does AtScale interact with my data platform?

AtScale acts as 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 cube is published, it is immediately available for BI and/or AI/ML activity. There is no pre-processing or data movement 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 end user's query tools and rewrites queries for execution on a data platform, leveraging any available AtScale managed aggregates that would be beneficial to the user's query.

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

What are the options for aggregate creation?

Aggregates may be triggered in 3 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 are stored inside the AtScale model. Users can specify combinations of dimensions and measures to design an aggregate manually and these aggregates will automatically build when the model is published.

In addition to these types, settings are available for adjusting behavior and thresholds for creating 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 on a file trigger basis by using AtScale's file watcher utility. This method is often used in conjunction with an ETL pipeline to trigger a refresh upon 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 the appending of new or changed data whereas a full refresh will rebuild the aggregates from scratch.

ATSCALE