

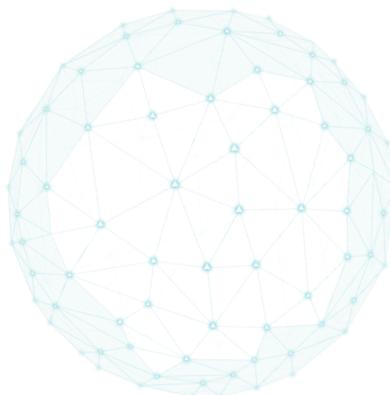
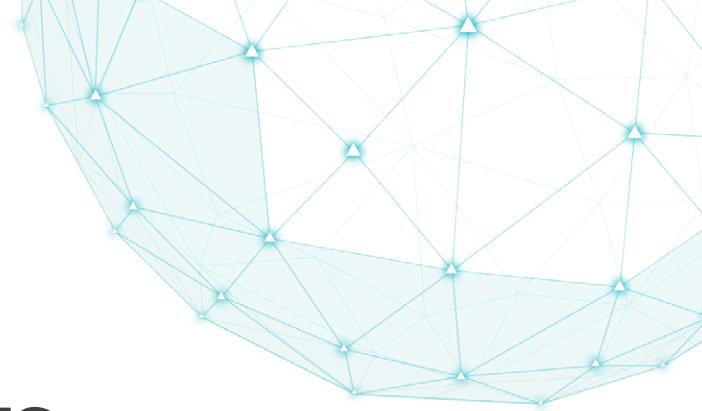
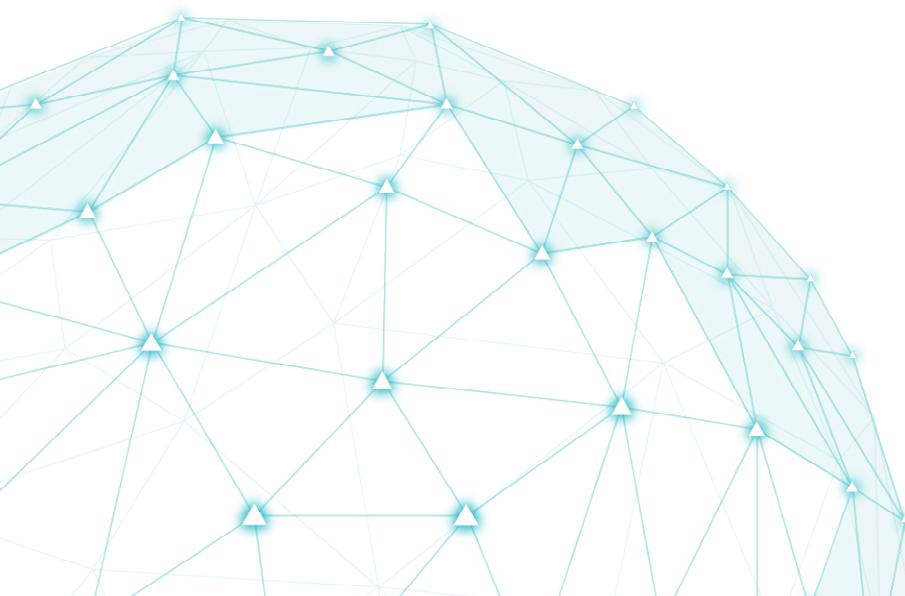


TECHNICAL OVERVIEW

September 2019

TABLE OF CONTENTS

Why AtScale?	1
System Overview	3
AtScale Universal Semantic Layer™	3
AtScale Accelerated Data Structures™	4
AtScale Design Center Canvas™	4
AtScale SDK	5
System Architecture	6
Overview	6
Virtual OLAP Cubes	7
Deployments	11
Integrations	11
FAQs	12





WHY ATSCALE?

Enterprise organizations are realizing that simply putting lots of diverse data into data platforms can't generate meaningful insights without further integrations, transformations, enrichment, and orchestration. Delivering connected data across on-premises and cloud sources is not for the faint of heart, especially with growing data volumes, complexity, and the need for a high rate of ingestion. Poorly integrated business data leads to inaccurate business decisions, reduces customer satisfaction and competitive advantage, and slows innovation and growth.

Organizations need to democratize data to allow business consumers to leverage and collaborate with all data to support faster and more accurate business decisions. Insight-driven enterprises require access to a complete set of data, leveraging self-service and real-time data architectures to deliver trusted, accurate and timely insights and shared data intellect.

AtScale is positioned to address this imperative by breaking down data silos, addressing data privacy regulations with stronger security measures, enabling real-time data sharing, and opening access to a complete and trusted set of data via self-service with large-scale data volumes on modern data platforms. AtScale achieves this uniquely with the following features and functions:

- ▲ **No data movement:** AtScale supports many modern data platforms, whether on-premises or in the cloud. Our Accelerated Data Structures are stored on disk alongside your source data, on your data platform. When installing and configuring AtScale, we require a schema on the platform where aggregates can be written. The aggregates are stored on the platform without additional movement.
- ▲ **Automatic ‘smart’ aggregate creation:** The AtScale platform includes a machine learning component that dynamically optimizes the physical data structures on disk based on your users’ actual behaviors—reducing or in some cases eliminating the need for traditional manual ETL or “data wrangling” which can take weeks or months. AtScale automates the data engineering to create and maintain optimal data structures to provide the best possible response time for your users, without any undue latency between when the need for an aggregate arises, and when that aggregate can become available for use. As time goes on, AtScale continuously and automatically adapts these optimized data structures in real time, constantly responding to any changes in user behavior and the models/data that you provide to your users.



- ▲ **Optimize the data and the queries:** Because the AtScale Universal Semantic Layer™ has all the intelligence around your physical table structures, calculations, business rules, etc. built-in, your BI tools will generate and submit very simple queries to AtScale. When it receives them, AtScale will dynamically rewrite the queries, not only to accommodate the actual nature of the data and the query engine in use, but also to make sure that any available aggregates are used that could accelerate the performance of the query. End users simply ask for the data they need, and AtScale makes sure the query to get it is as optimal as possible.
- ▲ **Use your existing BI & AI/ML tools:** BI and/or AI/ML tools need to work on data platforms and AtScale seamlessly provides that access without the user having to understand where or how to access them. Leverage what you have, what you know, and what you depend on.
- ▲ **No more extracts or shadow-IT systems:** Extracts in the form of .csv or .xls formats for Excel, .tde files for Tableau, and similar files that are proprietary to any given BI or AI/ML tool lack governance and controls. Moreover, extracts introduce questions about accuracy, validity, and the freshness of the data. Data can also be stored locally on a shadow database being fed and maintained outside of the governance and control of the data lake. AtScale provides a single source for a full and live dataset with validity and accuracy guaranteed.
- ▲ **Data as a service:** Data consumers need consistent answers to their queries and that's not possible when proprietary metadata is re-created for each tool. What's worse, when using tools without a metadata layer, the user is required to have a specific proficiency which breeds variations and human error. With AtScale, metadata is created once and includes the joins, business rules, calculations, etc. creating 'Data-as-a-Service' and ensuring the validity and accuracy of every query from every tool.
- ▲ **BI & AI/ML portability:** Once a model is built in AtScale, it becomes portable without the need to recreate it if moving from on-premises to the cloud. AtScale simply re-points to the new platform and handles everything on the backend such as changing the syntax of outgoing queries to automatically match up with the new data platform. Transitioning to the cloud couldn't be easier.
- ▲ **Time to result:** Most traditional data warehouse/mart environments require weeks-long processes to source, query and model data. With AtScale, lead times to get analytical projects into production drop from weeks to hours. Without the need to manually create, test, and maintain ETL jobs to do the aggregation of data, and with an intuitive UI, AtScale models can be created and put into production during the course of a single day.



SYSTEM OVERVIEW

AtScale's Adaptive Analytics Platform is an analytics fabric that removes the complexities of delivering agile, governed self-service advanced analytics to the enterprise, allowing you to shift resources from managing data to analyzing it. AtScale's autonomous data engineering creates, maintains and efficiently optimizes its acceleration data structures over time to alleviate distributed joins and determines where in the data fabric to store them for further optimization and efficiency. Our Universal Semantic Layer™ reduces data management complexities and gives users access to live data in a secure and accurate manner no matter its source. AtScale connects business intelligence and data science tools to on-premises and cloud data platforms in real time via the Accelerated Data Structures.

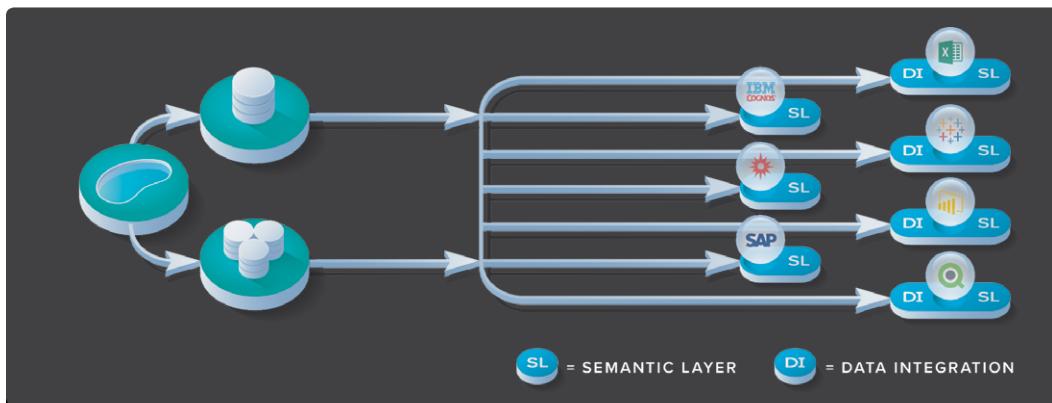
AtScale Universal Semantic Layer™

AtScale models your data as a virtual OLAP cube using a multidimensional OLAP cube metaphor including dimensions, hierarchies, measures, etc. AtScale creates a business-centric view of the data using a familiar modeling interface used in BI and AI/ML tools such as BusinessObjects Universes or Cognos Framework Manager. Using a familiar construct and design interface, users can quickly and easily source and model company data, incorporating table joins, business rules, calculations, etc.

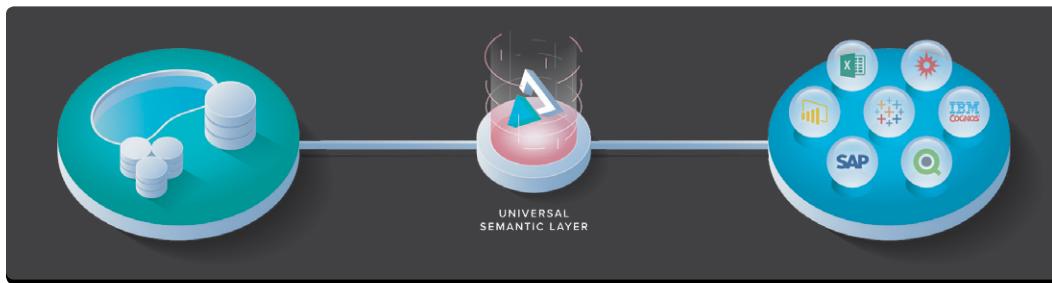
Once a model is created, it can be published via ODBC, JDBC, and MDX for BI and AI/ML tools to query against. There is no latency between model creation and when users can query against it. The Universal Semantic Layer allows any tool to connect to AtScale and access models without the need for users to join tables, create calculations or business rules, etc. AtScale unifies both data storage platforms and data analytics tools utilizing centralized management and governance across the board.



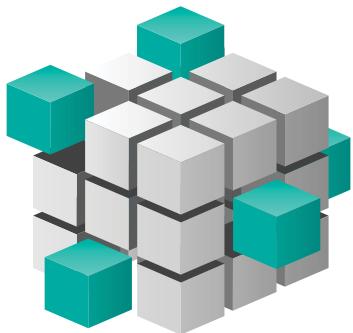
BEFORE ATSCALE



AFTER ATSCALE

**AtScale Accelerated Data Structures**

Deploying a modern data platform with large-scale data such as Hadoop, Microsoft Azure, Amazon Redshift, or Google BigQuery almost always finds issue with query performance. AtScale addresses this pervasive problem via machine learning. AtScale monitors, analyzes and optimizes data queries based on how data is being used, learning what data is important, and how best to shape that data into aggregates, providing conversational response time. Queries that run against many terabytes of data can yield responses in tenths of a second from an AtScale aggregate. Collectively, these aggregates are referred to as ‘Accelerated Data Structures’ because they constantly adapt to changes in user behavior.



With AtScale, the moment a model is published, users can access that data. The aggregates are built in real-time, in response to user activity, and automatically incorporated into user queries without any additional manual intervention. It's via these aggregates, which are stored on disk alongside original data in a chosen data platform, that virtual OLAP cubes are materialized based on the users' actual behavior.



AtScale Design Center Canvas™

The AtScale Design Center is an HTML 5 interface, accessed via a web browser and used by model developers to create and publish AtScale models (much like working with BusinessObjects Universes, Cognos Framework Manager, or SSAS design). The Design Center Canvas functions around the following structures:

- ▲ **Organizations:** The ‘top-level’ structure contains sets of users, groups, roles, permissions, etc. 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 structure under which one or more virtual OLAP cubes will be developed. Within a given project there is a shared library, in which the various objects are created (like Dimensions) and are stored for common re-use throughout the cubes in that project. There is no limit to how many Projects you can create with as many cubes within a given project as needed.
- ▲ **Cubes:** A Cube is created within a Project and can use shared objects from that Project’s library. A Cube is an individual curated virtual data set, containing the metadata the designer has specified, with all the joins, business rules, calculations, etc. defined and ready for use by data consumers with their BI and/or AI/ML tools.

AtScale enables BI and/or AI/ML tools to be used on any data platform, with hyper-scale volumes of data, and with the conversational response time data consumers need.

AtScale Data Platform SDK

In early 2019, AtScale released a data platform Standard Development Kit (SDK) to standardize the development of data warehouse support and enable 3rd party development. The new SDK enables users to more easily add support for new data platforms and extend coverage for native platform capabilities. The platform allows for the addition of support for new data sources without requiring an AtScale release to make it generally available.



SYSTEM ARCHITECTURE

Overview

AtScale installs on a single Linux server/VM (or multiple, for HA purposes) within any environment, either on-premises or in the cloud. From that instance data can be accessed and modeled in the AtScale Design Center (a browser-based HTML 5 interface) as a virtual multidimensional cube.

Once the model is ready for use, it is ‘published’ in AtScale which makes it available for querying from any BI and/or AI/ML tool. These tools can ‘talk’ to AtScale via SQL (ODBC/JDBC connection) or via MDX (OLE DB/ XMLA connection). When using ODBC/JDBC, AtScale’s engine is exposed as a Hive thrift service (without the need for a proprietary AtScale driver). Clients use a standard Hive ODBC/JDBC driver. No driver is necessary when using an MDX-based tool such as Excel. MDX is accommodated via the transport and the OLE DB spec itself. AtScale will receive incoming queries from a BI and/or AI/ML tool, and rewrite it to source data from the data platform as well as making use of available aggregates in the Accelerated Data Structures.

To BI and/or AI/ML tools, AtScale is a server. To a data platform, AtScale acts like a client.

As data consumers interact with the model, AtScale will automatically create aggregates through the Accelerated Data Structures. AtScale does not move data; data remains on the data platform.

AtScale will interact with the storage and query engine interfaces of any platform. For example, on AWS Redshift, AtScale can act as the engine and S3 via Spectrum to access data. On Hadoop, AtScale interacts with the Hive metastore to determine the location and structure of data in HDFS, and Hive, Spark, Impala, etc. as the SQL-on-Hadoop query engine. The capability and functionality of AtScale does not change regardless of the data platform choice or whether it’s on-premises or in the cloud.

NOTE: There is a REST API for AtScale which can be used to automate all aspects of the AtScale platform. Using the API you can programmatically automate AtScale for use in any of your application development or integration efforts.



Virtual OLAP Cubes

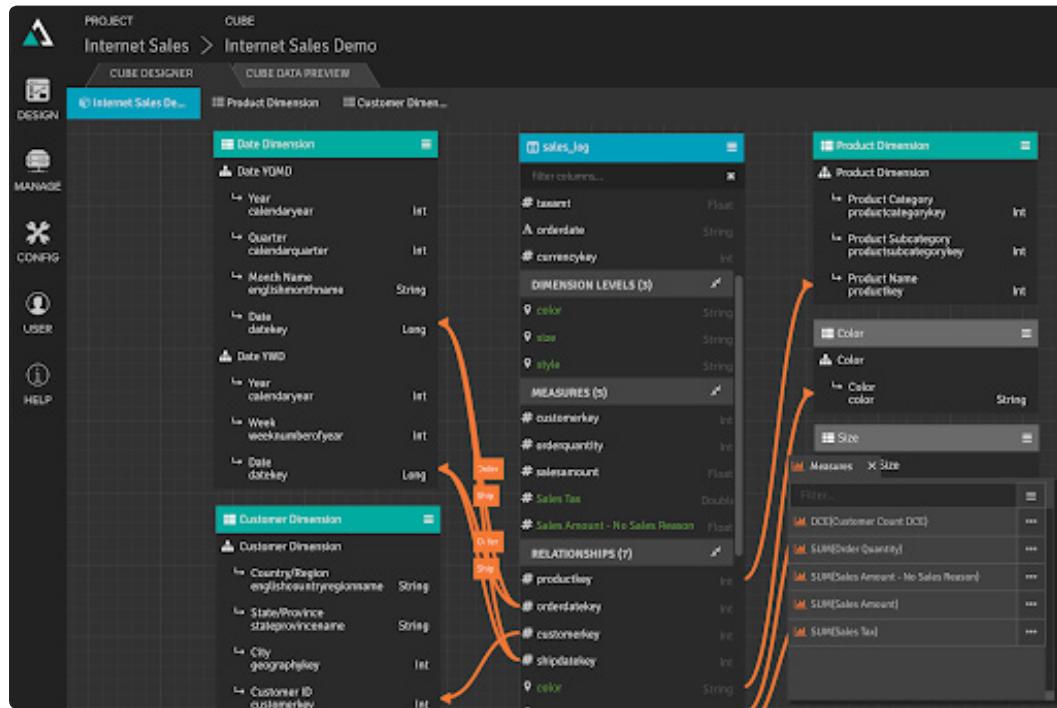
Development in the AtScale Design Center is done in the OLAP metaphor called Online Analytical Processing. AtScale provides a business-centric view of data, focusing not on technical aspects such as SQL or MDX code, but on arranging the data in a way that is intuitive and easy for end users to work with. As such, the different facets of data are considered ‘dimensions.’ Data is organized into dimensions to any BI and/or AI/ML tool. Measures (the facts or metrics, around which your dimensions create context) are created as well.

Models in AtScale visually look like a ‘star’ or ‘snowflake’ schema similar to what might be designed for a data mart. There is no dependency in AtScale for any particular physical data structure on disk (if data is in a normalized and/or data warehouse-style).

The screenshot shows the AtScale Design Center interface. The top navigation bar indicates the project is 'Internet Sales' and the cube is 'Internet Sales Demo'. The main workspace is titled 'Customer Dimension' and contains a grid of dimension and measure definitions. On the left, there's a sidebar with icons for Project, Design, Manage, Config, User, and Help. The central area shows a dimension named 'customer_file' with various attributes and relationships. A red arrow points from the 'customer_file' section towards the 'Hierarchies' pane on the right, which displays a hierarchy structure for the Customer Dimension, starting from Country/Region and branching down to City, Customer ID, Gender, and Occupation. The bottom of the screen shows a preview pane with some sample data.

A customer dimension built from 2 base tables specifying a hierarchy with Country/Region at the top, following with State/Province, City, and Customers. Customers have two additional dimensions attached to the Customer ID level format. There is no limit to logical model creation regardless of what the physical model is on disk.

Along with dimensions and measures, models will require joins between the various tables that are involved in that cube. These joins are typically based on key structures between tables. With dimensions, measures, and joins, models can be published and are ready for querying.

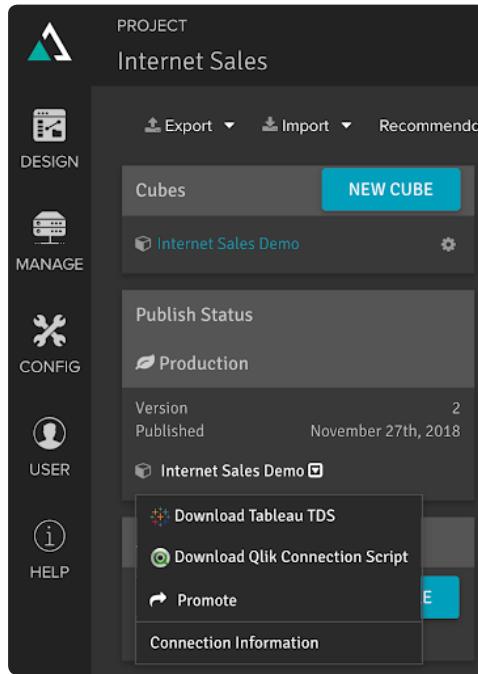


There are a number of measures coming from the 'sales_log' table, such as Order Quantity and Sales Amount. Items that are in green text are calculated items that have been created in AtScale and do not natively exist in the data on disk (such as Sales Tax). The series of orange lines connecting dimensions to the fact table denote where joins have been defined in the model.

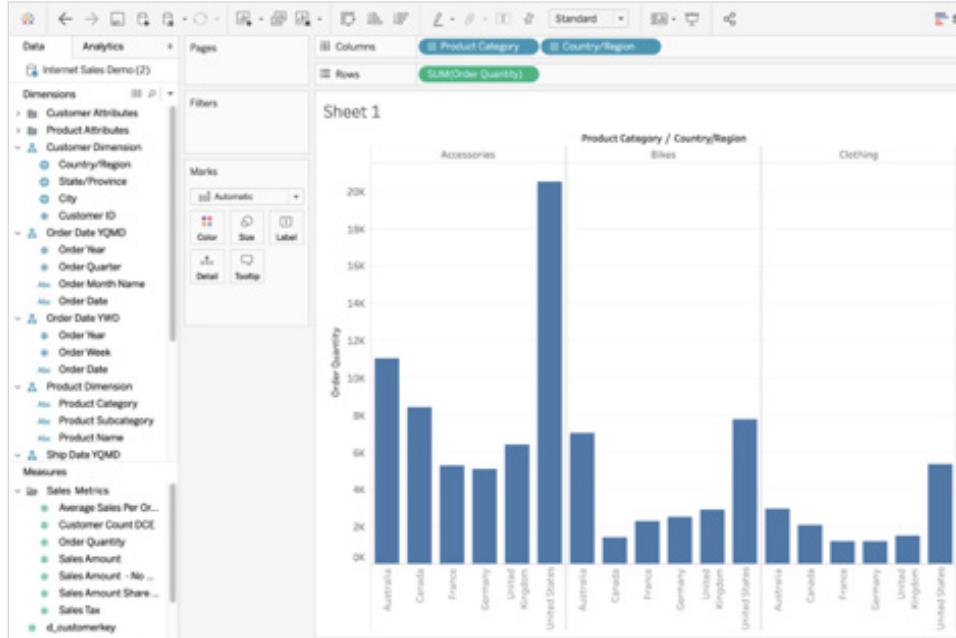
The screenshot shows the AtScale Integration interface for the 'Internet Sales Demo' cube. The left sidebar has options for DESIGN, MANAGE, CONFIG, and USER. The DESIGN tab is selected. The main area displays connection information:

Interface	Connection Information
JDBC	jdbc:hive2://master.corp.atscale.com:1111/internet+Sales
MDX	http://master.corp.atscale.com:10502/xmla/default/prod
ODBC & SQL	master.corp.atscale.com:1111

Once a model is published, the connection information is displayed - ODBC, JDBC, or MDX.12



Connection information can be automatically downloaded for tools, such as a Tableau .tds file which is the metadata that Tableau uses to connect to a data source. Once published, the AtScale cube can be accessed via BI and/or AI/ML tool of choice.



A sample view of a Tableau report with the dimensions and measures on the left having been inherited from the AtScale model without any additional work in Tableau (running on SQL.14).



The screenshot shows an Excel PivotTable with the following data:

	Order Quantity
Row Labels	Order Quantity
Australia	
Accessories	11032
Bikes	7032
Clothing	2952
Canada	
Accessories	8432
Bikes	1455
Clothing	2096
France	
Accessories	5299
Bikes	2304
Clothing	1211
Germany	
Accessories	5102
Bikes	2512
Clothing	1213
United Kingdom	
Accessories	6403
Bikes	2919
Clothing	1515
United States	
Accessories	20535
Bikes	7765
Clothing	5390
Grand Total	95167

The PivotTable Fields pane on the right shows the dimensions and measures used:

- Customer Dimension:
 - Customer Dimension
 - Country/Region
 - State/Province
 - City
 - Customer ID
 - Customer Attributes
 - More Fields
- Order Date Dimension

Drag Fields between areas below:

- Filters
- Columns
- Rows: Customer Dimension, Product Dimension
- Values: Order Quantity

A view of the same AtScale cube in Excel. The same dimensions and measures are here as in Tableau (running on MDX).

AtScale's Universal Semantic Layer provides the same logical interface to the end user regardless of the BI and/or AI/ML tool. Users can interact with data using the same dimensions, hierarchies, and measures using the calculations and business rules that have been included in the model. With AtScale, you democratize data and deliver it as a service or 'Data-as-a-Service' to all data consumers without any restrictions to share and collaborate—and with conversational response time as a function of the Accelerated Data Structures.

All data, to all people, at all times.



DEPLOYMENTS

AtScale is deployed on one or more Linux machines or VMs via a single RPM/DEB. A High Availability (HA) deployment is available via the same RPM based installation.

The AtScale Platform consists of multiple services:

- ▲ AtScale Design Center
- ▲ AtScale Engine
- ▲ AtScale Service Controller
- ▲ AtScale Coordinator
- ▲ Postgres

AtScale requires the following software to operate:

- ▲ Active Directory, LDAP, or cloud-based Identity platform
- ▲ Load Balancer

INTEGRATIONS

AtScale supports data platforms more than just SQL dialect translation. AtScale understands the semantics of the queries as well as the native platforms' underlying capabilities for security, performance and agility. As a result, AtScale can holistically evaluate many different strategies for query optimization and subsequent autonomous data engineering. AtScale supports Hadoop-based SQL engines, cloud-based data warehouses, traditional RDBMS, and BI and AI/ML tools.



FAQs

What do I need to deploy AtScale?

- ▲ A supported data platform, such as Hadoop (on-premises or in the cloud), Amazon Redshift, Snowflake, Cloudera, Google BigQuery, etc.
- ▲ A Linux server/VM with some basic prerequisites to install the AtScale software.
- ▲ An AtScale license and access either to the online marketplace for your chosen cloud platform (Microsoft, Amazon, or Google) or the AtScale installer downloaded for your Linux and/or Hadoop distribution.
- ▲ Appropriate JDBC/ODBC drivers for your BI tools. No driver necessary for MDX tools like Excel.

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

- ▲ AtScale supports a proof-of-concept trial. Please [contact us](#) and 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 a data platform and will interact with its storage and query engine to access the data available and model it as a virtual OLAP cube.
- ▲ 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 or MDX and begin querying the cube.
- ▲ AtScale receives these incoming queries from end users acting as a server to the BI and/or AI/ML tools. AtScale rewrites queries for execution on a data platform and leverages any available aggregates in the Accelerated Data Structures that would be beneficial to the user's query.
- ▲ Simultaneously, AtScale's machine learning algorithms are monitoring actual user activity with the data and responding by building out and evolving the Accelerated Data Structures to serve the actual needs of the end users. These aggregates are created and stored into a schema which is stored on your data platform.



Are there other options for aggregate creation in the Accelerated Data Structures?

- ▲ Yes. In addition to the demand-based, or heuristic, creation of aggregates based on users' query behaviors, there is the ability to enable 'Predictive Aggregates'. AtScale will analyze models at the time they are published and will proactively generate aggregates based on the design of the model.
- ▲ There is also an option to create 'User-Defined Aggregates' which can specify combinations of dimensions and measures to design an aggregate manually. Settings can be adjusted to control how the automatic, demand-based aggregate generation is managed.
- ▲ In a 'training mode' AtScale allows for pre-generation of demand-based aggregates without having to wait for users to interact with the model to trigger them. Once a cube is published, AtScale can be set in training mode where queries can be made without processing. AtScale will compute what it has learned and apply it to build demand-based aggregates based on that particular behavior pattern.

How are the Accelerated Data Structures managed and kept current?

- ▲ There are a number of settings that can be adjusted to control the creation of demand-based aggregates, as well as how much cache AtScale can maintain. AtScale has a built-in scheduler for cache maintenance—either time-based or set as a file watcher to trigger an update on existing aggregates. The API is also often used to integrate the AtScale aggregate maintenance into a given workflow.

NOTE: Cache can be updated either incrementally or as a rebuild. Incremental updates allow for the appending of net-new data to the existing data in the aggregates whereas a rebuild allows for adjustments to historical data as well as resetting partitions in the aggregates.

What does AtScale do to my data?

- ▲ AtScale does not change data stored within your data platform. AtScale queries against data but will not make any changes to source files. AtScale creates aggregates and stores them in a granted schema which does not impact source data files.

ATSCALE