



WhitePaper
2020

Achieve Data-Driven Insights with AtScale Cloud OLAP

Where Traditional OLAP Fails, Cloud OLAP Succeeds

ATSCALE



ENTERPRISES EXPECT TO PERFORM ANALYTICS IN THE CLOUD

[Ventana Research shows](#) that 86% of the organizations surveyed expect the majority of their data to be in the cloud and a whopping 99% expect to do their analytics in the cloud. One compelling reason that organizations are moving to the cloud en masse is to simplify their analytics stack. In the cloud, business users are freed from the burden of managing their own data platform clusters and can expand and contract their data resources to meet demand without the friction of dealing with traditional data center provisioning.

The New Cloud Analytics Stack

Hadoop was the beginning of the data lake concept and challenged the dominance of the highly structured, traditional enterprise data warehouse (EDW). Now, each of the public cloud vendors has their own version of a data lake. Their respective distributed file systems (AWS S3, Azure Synapse and Google Cloud Storage) have become the landing zone for data in the cloud. The rebirth of the data warehouse in the cloud with the likes of Snowflake, AWS Redshift, Azure SQL DW and Google BigQuery, provides the comfort of the EDW but with the elasticity and the operational ease of the cloud. However, even though there is no single repository for data in the cloud, enterprises are still striving for semantic consistency, data governance and the ability to manage their data in one central location. To achieve this goal, AtScale offers an Intelligent Data Virtualization layer which connects different data sources and makes them accessible from a common logical data access point for Cloud analytics.

QUICK HISTORY LESSON: ONLINE ANALYTICAL PROCESSING OR “OLAP”

OLAP is the acronym for OnLine Analytical Processing. Database researcher, E. F. Codd, coined the term “on-line analytical processing” (OLAP) in a whitepaper published in 1993.

When OLAP was created, databases were essentially two dimensional - records and fields - and required a query language (SQL) in order to retrieve data. With OLAP, users could use “cross tabs” and pivot tables to formulate their queries. Like a spreadsheet, OLAP allows for cell-based calculations and queries, presenting data in a multi-dimensional fashion instead of just rows and columns.



Business users could simply use a BI tool or Excel to drill down, pivot, and swap dimensions and measures. It quickly became the default language of business. Instead of looking at records, fields, and facts, any user could look at “Sales by Region”, “Sales by Product”, “Sales by Channel”, etc. These “by’s” define the actual multidimensionality of OLAP and the ability to just drag and drop, drill down, drill up, etc. with the “click of a mouse” instead of writing SQL query.

CLOUD OLAP SUCCEEDS WHERE TRADITIONAL OLAP FAILS

According to Gil Press, Senior Contributor to Forbes, in his [“A Very Short History of Big Data,”](#) the term “Big Data” started to get used in 1998. By 2001, Doug Laney, an analyst for the Meta Group, was starting to write about the three V’s: volume, velocity, and variety of Big Data. In 2002, Doug Cutting and Mike Cafarella were working on an Apache Nutch Project that involved building a web search engine that would crawl and index websites. This turned into an open source project called “Hadoop”. By 2008, Yahoo and Facebook started using Hadoop and Hadoop technology started defeating supercomputers to become the fastest system on the planet by sorting an entire terabyte of data.

AtScale co-founders, Dave Mariani and Sarah Gerweck, were at Yahoo at this time and they were using OLAP with Hadoop to analyze Yahoo’s browsing and advertising data. While they achieved their business objectives of delivering real value for consumers and promoting an amazing degree of self-service, the architecture proved to be too rigid and too fragile, didn’t scale, wasn’t securable and was not sustainable.

Through the Hadoop era, OLAP cubes were still widely in use and “exploding” with data. For some Big Data users, OLAP and “Cube” had become challenging. OLAP administrators and power users began feeling the pain as their data grew exponentially while their business users still expected the same “speed of thought” response times. The co-founders of AtScale were also feeling the pain and set out to create a [new type of OLAP](#) for modern data architectures in the Cloud.

As you can see in the chart below, the new OLAP model solves for many of the shortcomings of the traditional OLAP approach while preserving OLAP’s benefits of speed, simplicity, calculational power and semantic consistency.



REQUIREMENT	ATSCALE CLOUD OLAP	LEGACY OLAP	"DO IT YOURSELF" DATA MARTS
Ad hoc query performance under 1 second	Yes (Automatic)	Yes (Automatic)	No. Manual engineering required and sub-second response time is impossible.
Universal Semantic consistency & governance	Yes	Yes	No
Ability to handle complex business models	Yes	Only for Tools that speak MD	No. SQL is not suitable
Scale to large datasets with high cardinality	Yes	No (pre-computation/memory limited)	Yes. Manual data engineering required.
Predictable cloud costs	Yes	Yes for Azure but N/A for other clouds because they don't offer OLAP.	No. No way to control errant queries.
Quick response time to introduce new data	Yes	No. Must rebuild cube and change ETL.	No. Requires manual data engineering & ETL modifications.
Hybrid cloud support	Yes	No. Only Azure supports OLAP & other clouds do not.	No, except for Snowflake. Manual data engineering & ETL are cloud specific.
Unified security & authentication	Yes	No. Separate data store requires separate security.	No. Each BI tool requires different security settings.
Heterogeneous BI tool support w/ live connection	MDX & SQL	Only MDX	Only SQL
Avoids redundant data copies	Yes	No. Requires cube build.	It depends. Most likely requires different tables for different tools.

CLOUD OLAP (COLAP) IS THE FUTURE. THE FUTURE IS NOW

The [rebirth of the data warehouse](#) in the cloud with the likes of Snowflake, AWS Redshift and Google BigQuery, provide the comfort of the EDW with elasticity and operational ease of the cloud. Cloud OLAP (COLAP) makes OLAP work at cloud speed and scale to analyze large amounts of data without moving it out of the cloud data warehouse or data lake. Leveraging AtScale’s Cloud OLAP engine, business users have:

1. OLAP compatibility with Excel and any BI tool that speaks MDX or SQL
2. Fast, consistent and low-cost multidimensional queries
3. Simple modeling tools to instantly add new data elements and data sources
4. Direct access to any data warehouse or data lake -- including cloud data warehouses and nested data
5. Security and governance controls to manage data access in one place
6. A single source of truth for critical business metrics, defined server-side



Why AtScale Was Founded

By Dave Mariani, co-founder and Chief Strategy Officer, AtScale

While running data pipelines and analytics for Yahoo! in 2012, I struggled to keep up with the growth of our advertising data and make it usable for the business. We had truly big data before the term “Big Data” was invented. On top of the data volume challenge, I had several internal and external customers who all wanted to consume this data using different tools and applications. The result was utter chaos. Our most simple terms of “page views” and “ad impressions” had different definitions across our business units – there was no single source of truth for our most basic business metrics and my users didn’t trust our data.

I needed a Universal Semantic Layer, or data API that would:

1. Work with a variety of data consumers and tools
2. Deliver fast, consistent query performance
3. Define secure metrics in once place with an easy to use business interface
4. Provide a self-serve interface for business analysts and data scientists
5. Scale with our business
6. Minimize manual data engineering
7. Avoid data copies and data movement

At the time, there was no single solution that could meet my objectives. So, to satisfy these requirements, I forced a shotgun marriage of Hadoop, Oracle and SQL Server Analysis Services (SSAS). The data landed in Hadoop, was pre-aggregated and ETL’ed into Oracle as a staging area, and then processed into a SSAS cube for end-user consumption. My SSAS cube ended up becoming the largest SSAS cube in the world at 24TB for 3 months of data. While five times larger than the next largest cube, 24TB is really not “Big Data” in my definition. Yet, keeping that beast fed required a delicate dance involving NetApp snapshots and tricked out ETL code. But my end users loved it and they were able to do wonderful, revenue-producing things with it. I really needed a way of having my cake (OLAP functionality) and eating it too (without the OLAP architecture).



ATSCALE HAS YOU COVERED FOR CLOUD ANALYTICS

AtScale's Cloud OLAP Engine provides the cloud boost enterprises are looking for without the disruption caused by redesigning data models or eliminating existing BI and AI tools. AtScale brings the query speed and centralized business logic that BI professionals love with Cloud OLAP and leaves its underlying architecture (the things BI pros didn't love) behind. By taking the best of OLAP and combining it with the latest cloud technology, AtScale delivers a solution that doesn't involve data movement but offers a direct query architecture. It's the best of OLAP for the Cloud. It's COLAP and with it, you have:

- ▲ "Speed of thought" query performance without manually building aggregate tables or hand tuning queries and reports.
- ▲ Ability to handle critical business calculations that SQL cannot deliver but OLAP (MDX language support) can. This includes multi-level metrics, time intelligence and semi-additive measures.
- ▲ A Universal Semantic Layer™ for the cloud data warehouse that supports multiple BI and AI tools with business logic and definitions defined in one place, server side.

Ready to try AtScale? Request a demo at: www.atscale.com/demo.

ABOUT ATSCALE

The Global 2000 relies on AtScale – the intelligent data virtualization company – to provide a single, secured and governed workspace for distributed data. The combination of the company's Cloud OLAP Engine, Autonomous Data Engineering™ and Universal Semantic Layer™ powers business intelligence and machine learning resulting in faster, more accurate business decisions at scale. For more information, visit www.atscale.com.