

The logo for ATSCALE, featuring the letters A, T, S, C, and L in a dark grey sans-serif font. The letter 'A' is replaced by a teal triangle pointing upwards. The letter 'L' is replaced by a teal triangle pointing to the right. The background of the page features abstract geometric patterns of teal lines and triangles forming spheres and hemispheres.

# Cloud Transformation: The Next Virtualized Data Frontier for BI, ML, and AI

How intelligent data virtualization will unlock  
your enterprise data potential



Cloud transformation is happening. The shift to the cloud will drive \$1.3T in IT spending by 2022 (Gartner) and spending on public cloud services will total \$277B by 2021, a compound annual growth rate (CAGR) of 21.9% (IDC). Enterprises are migrating, lifting, shifting and re-engineering to seize the promised benefits of the cloud, with the hopes that, at the end of the migration, they will realize cost savings, increased agility and enhanced revenues. The last mile of cloud transformation is the process of making data useful and accessible through a process called cloud data transformation.

## **The Chasm in Cloud Data Transformation**

Cloud data transformation happens when enterprises seek to make as much data available to as many users as possible to improve alignment between workers and business units, increase the rate of insight and discovery and to accelerate the rate of innovation and revenue growth. More succinctly, cloud data transformation is about maximizing company data to grow the business.

**Cloud data transformation is about maximizing company data to grow the business.**

The size and complexity of enterprise data ecosystems means that cloud data transformation presents many challenges. Cloud data transformation involves:

1. Cloud data migration - data will always be moving. As enterprises retire legacy systems or move data between systems to reduce costs or leverage new benefits, they require solutions that minimize, or preferably alleviate, business disruption during migrations.
2. Data accessibility - the hybrid cloud is here to stay for the foreseeable future. Enterprises require solutions that level the playing field for data sources, even in circumstances where the data platforms are less performant or scalable.
3. Preserving and merging security - intricate security and governance rules have been engineered into each data platform. Enterprises require solutions to preserve the security of unique individuals, data platforms, databases, etc. and extend enforcement throughout the entire solution (data platform, database, acceleration data structures, etc.).
4. Uniform data analysis - users will have preferred business intelligence (BI) tools. Cloud data transformation solutions must assure uniform performance and equivalent results across many applications.



Hidden costs and insufficient interoperability may render those hopes for efficiencies into mirages for many enterprises. Seven hurdles stand in the way of organizations realizing concrete value for their cloud data transformation efforts:

1. Cloud platforms may require restructuring of the enterprise's data, forcing huge ETL and data translation projects to prepare data for ingestion.
2. Many vendors store data in proprietary formats, effectively chaining customers to one solution.
3. Some siloed or on-premises data is attached to legacy systems that cannot be moved without re-engineering those systems.
4. Interoperability with outside systems and BI tools is limited.
5. Security and entitlements are difficult to maintain when merging data from many silos that may have different users and configurations.
6. Limited strategic plans to realize business benefits beyond moving to the cloud, such as broadening user access or leveraging integrated data.
7. Hidden cost implications of cloud transformation often exceed initially anticipated or budgeted numbers.

These challenges stem from most cloud solutions being technical solutions to technical problems. They solve niche challenges such as query performance, data visualization or expensive storage. Individual cloud solutions save some costs and may add benefits that were not possible with data silos, legacy systems or on-premises solutions. Enterprises require a solution that serves not just technical needs, but the larger business objective, which is to accelerate increases in revenue with greater efficiency. Enterprises need a solution to enable access to clean, comprehensive data that enables users to act in support of common goals and derive better insights that improve company profitability.

**Enterprises need a solution to drive accessibility to clean, comprehensive data.**

**Enable users within the enterprise to act in support of common goals and derive better insights that will improve company profitability.**

Such a solution is the virtual data warehouse.



## VDWs: The Missing Bridge in Cloud Transformation

Enterprises will migrate to the cloud for two critical reasons: to reduce costs or to increase revenues and as a general rule will fund projects contributing to revenue generation and growth over projects that contain costs. When migrating to the cloud, if the inadvertent consequences are vendor lock-in and limited interoperability with a small number of BI tools, enterprises will find themselves left with the same challenges they had before, and a renewed need to migrate between systems in the future. To future-proof data access and realize data value more effectively, enterprises need Virtual Data Warehouses (VDWs).

VDWs are solutions that facilitate analytical interactions between data and the tools that consume data, allowing one or more data sources to be accessed securely and consistently by one or more software applications or BI tools. The ultimate goal is uniform and shared access to all data that is cost-effective, highly performant and secure—irrespective of where it is physically stored. Additionally, integrating and conforming with the data management, governance and security practices that the enterprise IT team meticulously built on-premises is of critical importance. The ultimate benefits are dramatically reduced costs for running data infrastructures, improved query performance and significantly deeper insights across the organization. To achieve this, the VDW must be agnostic to the analytical tools used and location of data sources, providing users with uniform and transparent access to data with any tool they choose.

VDWs unify corporate data and simplify access to users. They obfuscate the systems involved in the back end, freeing users to concentrate on working with data rather than learning complex data engineering methods, hunting for accurate data in forgotten silos or creating extracts of data that increase risk to security, integrity and reliability.



# The Four Cloud Data Transformation Capabilities of VDWs

VDWs have the following four capabilities when bridging the gap between users and data:

## 1 Bring all data (silos) to users' fingertips

IT organizations today have common server orchestration tools to manage their on-premises and cloud server resources. Universal data orchestration tools are needed to automate data engineering while providing a comprehensive view of analytical data. A VDW must provide users with uniform visibility to all the data in an enterprise's diverse analytical repositories. Regardless of the location, configuration, formatting or technology storing the data, the VDW will present data in a fashion that is immediately ready to query and combine with other data. Users need only connect to the VDW to access any analytical data they have entitlements to work with.

VDWs provide a kind of “hybrid intelligence” that brings portability and flexibility for data sources. If a siloed data source must remain in place due to a legacy system, users can still access it. If an on-premises data source is migrated to the cloud, users can still access it. Virtualization smooths out the bumps in migration to the cloud and even between clouds because the location of the data source no longer limits the end user.

## 2 Manage security and data governance across all users and data sources

The enterprise must preserve information about which databases are the “system of record” or maintain the authoritative record for each element of data, what changes were made and by whom. A VDW must have the ability to respect the entitlements of individual users on different data sources and carry those entitlements through to data structures that are materialized from a larger corpus of data or combined from multiple sources. Further, entitlements must be enforced within the analytical applications users leverage to access data through a connection pool. Furthermore, a VDW should have mechanisms for promoting data quality and managing how data is joined when working with multiple fact tables to ensure like data is matched appropriately.



The immense advantage of VDWs for security and data governance is that all queries and results pass through the VDW. Rather than a many to many relationship between users and data, enterprises using VDWs will see a many to one to many relationship. This enables the VDW to act as an enforcer of security and data governance. This quality of VDWs enables enterprises to leverage VDWs to enforce security rules and data governance more effectively.

### **3 Represent multiple data sources as one**

Some VDWs will go one step beyond basic data virtualization and provide a single, comprehensive view of the enterprise's underlying data platforms, presenting a “single table” view of data for consumption by BI and AI / ML applications. Enterprise data warehouse virtualization further breaks down data silos, makes data easier to understand as a unified entity and fosters more comprehensive analysis.

### **4 Interface to any major BI tool**

Business users, analysts and data scientists will come from disparate backgrounds and will have individual preferences for the BI tools they use when working with data. Rather than struggling to bend all users to a single standard for BI software, a VDW ensures that queries will return consistent answers across different tools. BI tools vary in how they query data, often due to different dialects of query language such as MDX, MySQL, PostgreSQL and so on, so VDWs will need to support query translation for many solutions. The VDW should promote flexibility to use the preferred BI tool of each user so the enterprise can focus on the data.

## **Example 1 - Migrating Data from On-Premises to Cloud Without Disruption**

A global home improvement retailer needed to modernize its Hadoop solution, in use by 4,000 stores. Hadoop had proven more challenging to manage and too expensive, so they opted to move to the cloud with Google BigQuery (GBQ). Also, users were comfortable with Microsoft Excel, and IT did not want to introduce a new BI tool, so any new solution would need to work well with Excel.

The migration from Hadoop to GBQ was completed over a weekend without any disruption to the business users. On Monday morning users were able to query the same reports without noticing any difference



other than faster response times. The VDW empowered the retailer by providing fast, easy access to 2.5 years of data, refreshed every 20 minutes through its platform in conjunction with Excel and GBQ. Store managers could now not only see their store, but could also benchmark against other stores within and across regions through the VDW. Users retained the ability to use Excel and avoided costly re-training efforts.

## Example 2 - Democratizing Access to Global Users

A global chemical producer was looking to give over 15,000 employees access to business data that was centralized, organized and accurate. The company generated petabytes of data that was getting siloed in disparate databases. BI tool users wishing to analyze data across multiple sources were forced to download copies of data and create their own local databases. The result was millions of downloads annually into local storage. The company needed a way for employees to work with disparate data sets, minimize extracts from central data repositories to local machines and optimize the quality and accuracy of data.

A VDW was used to connect disparate data together and model that data into virtual cubes that employees could connect to using their preferred BI tool. Employees could now find trusted data in a central, globally-accessible location. The VDW's ability to connect disparate data in a data lake to form virtual cubes created a single source of truth that could be used online by employees globally. The VDW has driven huge efficiencies and has eliminated millions of redundant, divergent copies of data. The near elimination of database extracts greatly reduced security threats related to the copying and storing data on local networks and computers.

## Maximize the Value Extracted from Data

All cloud migration efforts must serve business drivers such as reducing costs or increasing revenue. It can be easy to become distracted by the short-term technical benefits of cloud solutions, but to accept such distraction would inevitably cause the enterprise to stop short of realizing the ultimate benefits of the cloud.



A solution that helps enterprises de-emphasize technology so that business users can have simple, intuitive access to as much accurate data as possible, such as a VDW, is the best solution for far-seeing companies. Enterprises leveraging VDWs will be able to remove all the technical hurdles of disparate data sources and focus employee effort not on tending the garden, but on maximizing the value extracted from the company's data and enjoying the fruits of their labor.

## The AtScale Solution for Virtual Data Warehousing

AtScale enables on the four important capabilities of VDWs:

1. Brings all data sources forward to be visible and valuable to users.
2. Facilitates data security and data governance.
3. Allows disparate data silos to be viewed as a single data source.
4. Makes it simple and reliable to use any BI tool to work with data.

### 1 Brings all data sources forward to be visible and valuable to users

AtScale enables users to access data repositories regardless of the location or platform on which they reside. Whether the database is on an on-premises Hadoop cluster, Google BigQuery, Amazon RedShift, Snowflake or another system, users can find any data source connected to AtScale through the AtScale Design Center and add it to the model they are building. AtScale manages all translations of queries invisibly in the backend, and enables users to conduct joins between different fact tables so that users can assemble richer pictures from which to draw insights.

In addition to bringing all data into the forefront, AtScale also applies machine learning to create intelligent data structures that accelerate queries and reduce query costs. Users running queries on models generated with AtScale will find results will be many times, sometimes orders of magnitude faster. With the AtScale Adaptive Cache, on average, customers see 5 to 20 times faster queries with reduced load and drastically improved user concurrency.





**With the AtScale Adaptive Cache, on average, customers see 5 to 20 times faster queries with reduced load and drastically improved user concurrency.**

## **2 Facilitates data security and data governance**

AtScale supports best-of-breed security practices wherever possible, from End-to-End TLS protecting data transmissions, to LDAPS, Active Directory, IdP and SAML for authentication. To facilitate security with user and API access, AtScale supports JWT, CORS and REST—all current state of the art in their domains. The Kerberos standard is used internally for AtScale’s ticketing system.

AtScale will always begin at the source database to obtain security entitlement information for a query. Using a patent-pending technology called True Delegation, AtScale checks the security policies of the origin database each time a query is made and applies the entitlements and security policies of the origin database to any data structures called by the query. AtScale therefore ensures every query executed on the VDW is associated with the end user who generated the query. There is never any security leakage from synchronization lags since the parent database is always checked. In addition, security permissions are maintained in data structures and query results transparently, eliminating accidental exposure of restricted or private data.

AtScale’s virtualization capabilities also enable AtScale to aid in data governance. Virtualization eliminates the need to extract data for local manipulation, thereby protecting data from going stale or getting out of synch with the live database. AtScale’s ability to join fact tables from disparate sources online also aids data governance due to rules built into the AtScale Design Center that regulate how fields from different sources should be matched together, reducing the chance of incorrect or misleading results.

## **3 Allows disparate data silos to be viewed as a single data source**

AtScale’s virtualization also enables customers to treat all of their enterprise’s databases as a single data source and a single source of truth, greatly simplifying the process of finding and manipulating data. This aspect of AtScale’s virtualization extends the VDW’s ability to help enterprises maximize access to their data and thereby realize the value that can be extracted from that data.



#### **4 Makes it simple and reliable to use any BI tool to work with data**

AtScale utilizes a translation engine called the Universal Semantic Layer (USL) that acts as a universal translator between BI tools and the underlying data. The USL understands the interfaces and query dialects of the BI tools and of the underlying databases. When a query is made, the USL ensures that the result from a Microsoft PowerBI query would return the same results as the same query executed from Tableau or Microsoft Excel.

## **Embrace a Superior Way to Use Data**

Enterprises migrating to the cloud must go beyond simply “lifting and shifting” data into the cloud and embrace an approach that fundamentally transforms how data will be used. By abstracting data infrastructure, unifying security and data governance and providing flexible and consistent access to data, enterprises will be able to leverage superior insights and coordination that will directly support the business’s bottom line.



ATSCALE

