DATOMETRY

Datometry Hyper-Q

Contents

Overview

Intelligent abstraction has reshaped every element of the IT infrastructure in the past decade: from compute over storage to networking, every component of the stack has been disrupted by virtualization. Virtualization is synonymous with flexibility, interoperability, and granular control which translates into more choices and vastly more efficient operations.

Datometry redefines data warehousing by extending the principles and benefits of virtualization to the database stack. The Datometry Hyper-Q™ platform abstracts the interface to the database and lets applications once written, or customized, for an on-premises data warehouse run natively on any major cloud data warehouse—without requiring changes to code or configuration.

Through its innovation, Datometry fundamentally transforms how enterprises view applications. Applications are no longer an obstacle to cloud adoption as they can now be replatformed rapidly and with next to no effort. Also, the lifecycle of an application is distinctly prolonged and the intrinsic value of an application can be preserved while quickly adopting new database technology.

Sitting between applications and data warehouse, Hyper-Q makes applications database-agnostic. Hyper-Q continuously intercepts the application's communication with the database and translates and redirects it to the new cloud data warehouse. Hyper-Q intelligently translates, transforms, and emulates SQL statements and features like stored procedures, macros or recursive queries in real-time. Business intelligence, reporting, ETL and home-grown applications written and optimized for legacy data warehouses, can now run natively, instantly, and fully transparently on the new cloud data warehouse. Enterprises now have a radically different way to adopt cloud data warehouses with Datometry Adaptive Data Virtualization™ at a fraction of cost, time, and risk of conventional data warehouse migrations.

Datometry Hyper-Q Architecture

Hyper-Q takes the source SQL sent by the applications to the legacy data warehouse and generates the optimal target SQL for the cloud data warehouse required to produce bitwise identical results expected by the application. There are many useful, yet non-standard features and keywords present in the source SQL that are unsupported in the new cloud data warehouses (and, in fact, some of them like SET table may never be supported). The virtualization layer intelligently analyzes the queries similar to a database engine parser to determine whether to invoke translation, transformation, or emulation.

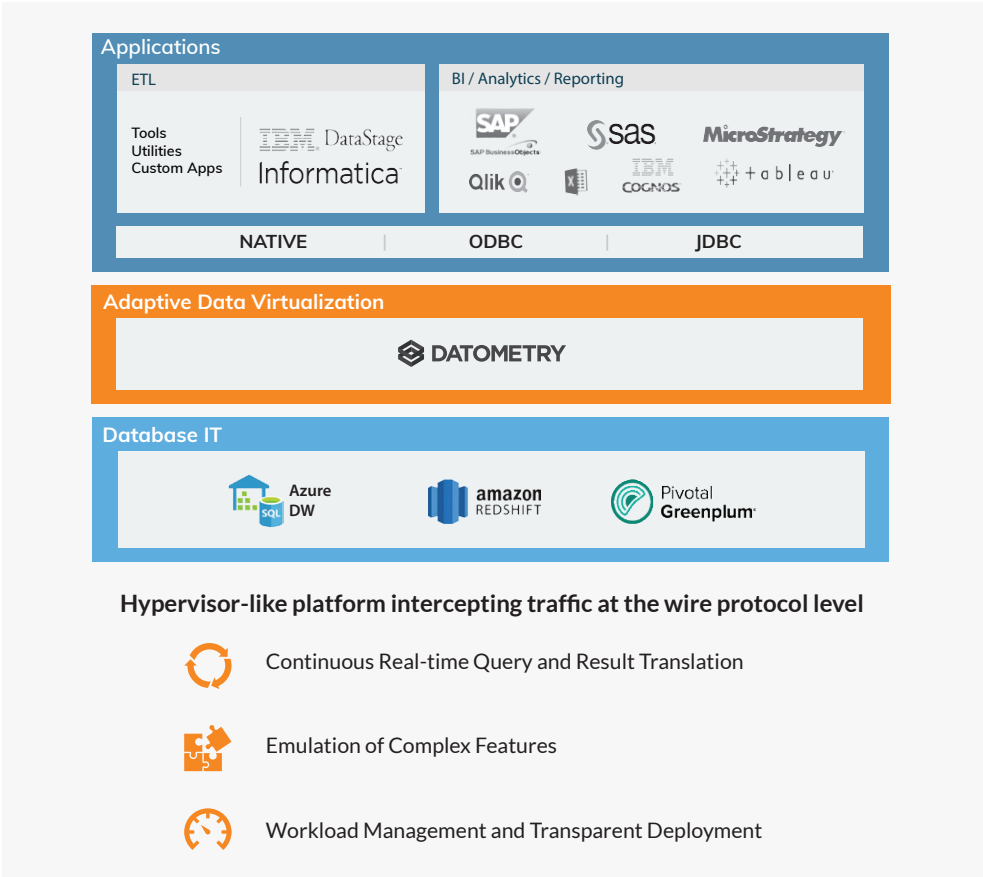
Hyper-Q's modular architecture allows for horizontal scaling, low-latency execution, and rapid

addition of SQL language extensions. In addition, its small footprint and transparent deployment (Figure 1) between the application and the cloud data warehouse means a fast and simple deployment. The transparent deployment is achieved by intercepting and manipulating application communication on the wire, at the transport layer. Hyper-Q requires no changes to the applications or the database and no agents on the application or the database. This stateless approach allows Hyper-Q to support a wide range of applications that connect through JDBC/ODBC drivers or the native protocol of the legacy data warehouse. Hyper-Q also performs workload management to help manage the concurrent query execution requirement of the applications.

At the heart of Hyper-Q is XTRA™, a powerful extensible Relational Algebra-based representation of queries and database commands. XTRA is used not only to model queries and commands, but also serves as the foundation for an in-depth analysis of requests as well as target-specific optimizations that enable Hyper-Q to play to each system's strengths. XTRA representation is then used to generate the specific cloud data warehouse SQL. This method of decoupling the source SQL parsing and target SQL generation allows for support of new source SQL, new cloud data warehouses, and support for vendor SQL language extensions (ex. QUALIFY) and variants (ex. date/time formats) in a very short period.

When processing requests, Hyper-Q retrieves catalog data from the server-side system to analyze the incoming requests and resolves references to database objects including tables, functions and views. Hyper-Q then performs full type derivation or inference and models the request at the level needed to perform semantic optimizations. Once optimized, XTRA is synthesized into language of the server-side system and submitted for processing through an ODBC connection.

Figure 1. Datometry Hyper-Q: Adaptive Data Virtualization



Continuous Real-time Translation: Query Translation and Query Transformation

Hyper-Q Translation is the simple syntactic replacement of source SQL to the target SQL; for example (Figure 2), the SEL keyword is used in some legacy data warehouse systems and can be directly replaced with SELECT to achieve the same results and keep the query fully compatible with the target data warehouse system.

SQL features, such as QUALIFY or MERGE, that are not fully supported today on cloud data warehouses are handled via Hyper-Q Transformation. Hyper-Q transforms (Figure 2) these keywords in source SQL statements and uses the currently available SQL features or keywords of the target cloud data warehouse to achieve the same results. GROUP BY ordinal position, date arithmetic, name resolution and implicit joins are also handled by the Transformation method.

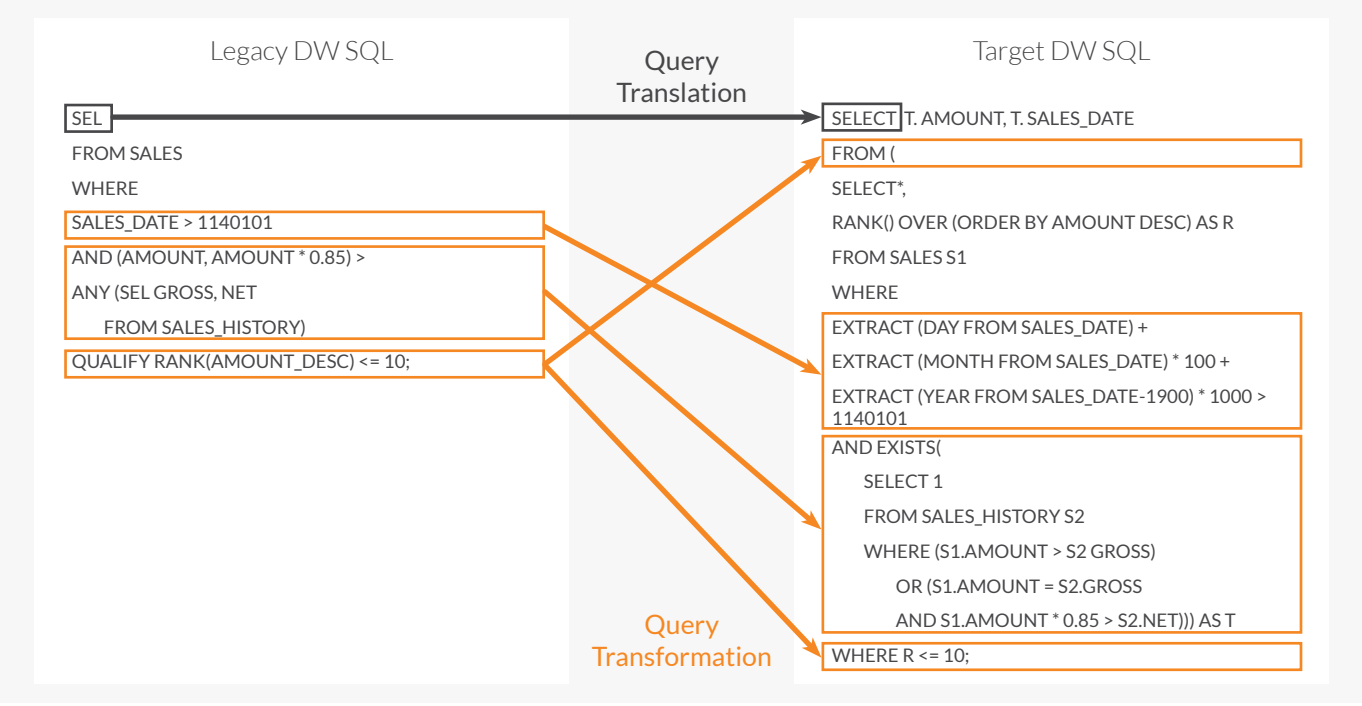
Augmentation of Missing Features: Full Feature Emulation

Advanced data warehouse features, such as stored procedures, macros, updatable views, catalog lookups and recursive queries, do not have identical equivalents in the cloud data warehouses today. In such situations, Hyper-Q provides full emulation of the missing features of the cloud data warehouse by using the available constructs on the cloud data warehouse.

Hyper-Q fetches the SQL statements within the stored procedures and macros, maintains execution state and variable scope, breaks out the control flow and executes the statements on the cloud data warehouse using temp tables as needed. The contents of the original stored procedure is preserved during the data/schema migration process using Datometry qShift™.

Recursive calls can become a nightmare for conventional replatforming. Hyper-Q breaks down recursive calls (Figure 3) constructs the call stack and uses temporary tables created on the cloud data warehouse to store the results at each step. Finally, when the recursion ends, the temporary tables are dropped, and the results are returned to the application in the legacy format.

Figure 2. Hyper-Q: Query Translation and Transformation



Updatable views and cascaded views are fully supported by Hyper-Q by retrieving the original legacy data warehouse view and dynamically building the map on each access by the application.

Workload Management: Query Concurrency and Connection Management

Hyper-Q provides sophisticated workload management facilities which ensure that workloads are not only fully functional but can be executed with the same operational profiles as expected from the original legacy system. Workloads can be managed using policy-based queuing determined by a variety of workload attributes including Teradata query bands. A rich set of management parameters, such as configurable timeouts based on query, connection, or session properties, give administrators control.

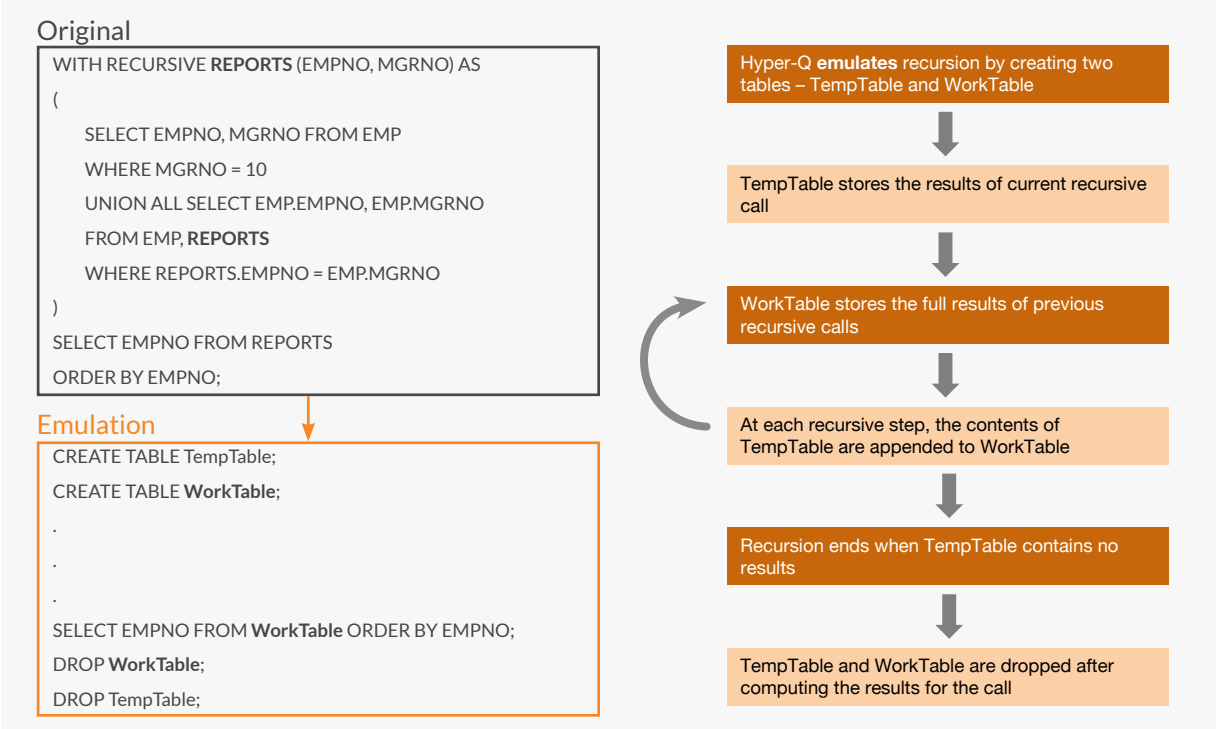
Finally, surge protection prevents overload situations during connection storms without compromising performance or functionality of existing connections.

Supported Applications and Systems

The client-server architecture of database systems defines APIs between application and database in the form of network communication. Hyper-Q provides applications with an API that is bit-identical to that of existing data warehouse systems. Applications can communicate with Hyper-Q in the same way as with the original database using the native wire protocol. Client applications do not need to be adjusted, rewritten, or reconfigured.

Connectors and drivers underlying ODBC or JDBC APIs as well as native communication libraries are supported and fully functional to provide support for a wide range of ETL, reporting, BI/analytics applications. Standard ETL tools, such as Informatica and IBM DataStage, reporting and BI applications like SAP Business Objects, SAS Analytics, MicroStrategy, Qlik,

Figure 3. Hyper-Q: Emulating Recursion



and Excel, have tested and are fully supported. Parallel loaders and utilities using native communication are also supported.

Deployment

Deployment	Description
SaaS Offerings	Hyper-Q is a SaaS offering for Amazon AWS and Microsoft Azure customers. It is deployed as software for Greenplum installations within a public or private cloud.
Scaling	Hyper-Q is specifically designed to scale horizontally with load balancers. A single Hyper-Q instance can manage 1,000s of active connections. In addition, Hyper-Q scales linearly with a simple load balancer distributing load to multiple instances of Hyper-Q due to its stateless nature. Hyper-Q bandwidth exceeds that of cloud databases by at least 10X since Hyper-Q does not maintain any state across SQL statements. All of the required metadata for name resolution and emulation used by Hyper-Q is stored on the cloud database.
Security	Hyper-Q SaaS is deployed within the private network of the customer within the cloud. Hyper-Q integrates with all common IT security standards and uses the same connection authentication mechanism as the underlying new database. Hyper-Q does not maintain any user data at rest, that is, Hyper-Q does not store any sensitive data. Both client-side and server-side communication can be encrypted using the legacy client applications and new database's supported encryption standards.
High Availability	Hyper-Q SaaS utilizes multiple Hyper-Q instances with cloud load balancers to match the 99.9% high availability of the cloud data warehouses.

Performance

Performance	Description
Latency	The overhead of Hyper-Q is negligible since it only translates statements and delegates the execution to the new database. In a benchmark test run, the Hyper-Q overhead was found to be less than 2%. Hyper-Q mainly manipulates and synthesizes SQL queries and does not generate results or perform complex operations on the result set like a database. The reduced number of operations in comparison to the database results in the very low overhead during the query execution.
Throughput and Concurrency	Hyper-Q is built using carrier-grade switching technology, which is purpose-built to handle massive numbers of concurrent connections with extreme throughput and stability requirements, resulting in negligible impact on system performance. The advanced workload management features of Hyper-Q specifically help with the optimal utilization of the cloud data warehouse. In some cases, Hyper-Q is able to increase throughput by batching multiple singleton inserts into bulk operations.

Operations

Hyper-Q is a SaaS offering that requires very few management cycles from an operational perspective since there are very few configuration settings needed for initial setup. There is no requirement to tune/optimize Hyper-Q instances.

Operations	Description
Monitoring	Hyper-Q is built on standard CentOS and integrates into most cloud monitoring dashboards.
Logging and Visibility	Hyper-Q provides maximum resolution of the SQL queries down to the packet level details. Both original and synthesized statement or statements are logged to allow for auditing and managing workloads. The Hyper-Q logs can be exported to an external log aggregation system.

Summary

Datometry enables data warehouse modernization and helps drive growth by providing the fastest path to the cloud whil lowering the TCO of the enterprise data warehouse by up to 80%. Unlike risky, time-consuming application rewrites, replatforming with Datometry takes only weeks and is significantly less risky as the business logic remains unchanged. This virtualization layer enables faster innovation and provides a shorter time to realize the benefits of the cloud.

The Datometry SaaS offerings of Hyper-Q, Datometry qInsight™, and Datometry qShift have enabled Global 2000 enterprises to replatform from Teradata® to modern cloud data warehouses, such as Microsoft Azure SQL Data Warehouse or Amazon Redshift, in weeks, not years.

Visit https://datometry.com to learn more about the enterprise replatforming journey to the cloud.

TBHQ Rev 10.0/18