ESG Economic Validation

The Economic Advantages of Google BigQuery versus Alternative Cloud-based EDW Solutions

By Aviv Kaufmann, Senior Validation Analyst

September 2019

Executive Summary

Enterprise data warehouses (EDWs) provide businesses with a wealth of actionable intelligence but are extremely costly to purchase and complex to manage, maintain, refresh, and scale. The needs of the business often quickly outgrow the capabilities of the system, presenting a challenge for admins and limiting the productivity of data scientists. Cloud-based EDW solutions are easier for organizations to maintain and manage as well as agile enough to meet the needs of the business. Cloud-based solutions such as Google BigQuery, AWS Redshift, Azure SQL Data Warehouse, and Snowflake can provide significant savings and benefits over an on-premises EDW, but they are designed and implemented quite differently.

ESG created a total-cost-of-ownership (TCO) model that compared the expected costs and benefits of satisfying the needs of a modeled organization with each of these cloud-based EDW solutions over a three-year period. Many of the costs and assumptions used in this modeled scenario were validated with real-world customers or by examining vendor case studies. ESG conservative models found that BigQuery can provide a three-year TCO that is 26-34% lower than the other three cloud-based solutions. The elimination of upfront investment, reduction in management complexity, and ability to integrate natively with other solutions contributed largely to the decrease. Moreover, ESG saw that the underlying architecture of Google BigQuery, decoupling processing capability and storage capacity, also contributed to lowering overall expenses.

Introduction

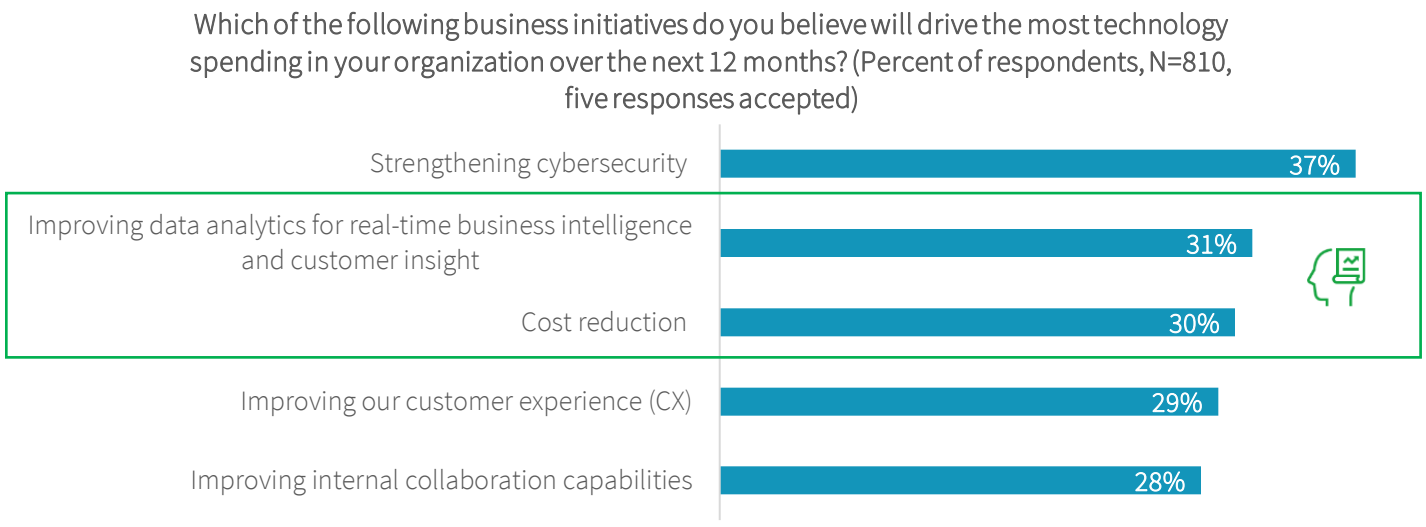
This ESG Economic Validation focused on the quantitative and qualitative benefits organizations can expect to realize with Google BigQuery when compared with other leading cloud-based enterprise data warehouse (EDW) services.

Challenges

The on-premises enterprise data warehouse (EDW) has been the backbone of many top enterprises over the past few decades. Organizations that possess the scale and IT budget to operate an at-scale EDW have reaped the tangible business advantages provided through timely and well-correlated intelligence, making the extraordinary expense well worth the large capital and operational investment. Today, organizations are collecting data from a much larger variety of devices, and data is growing at a much faster velocity. Monthly reports generated from legacy data sources often provide a historical data point rather than competitive advantage and intelligence. It is no longer enough to simply correlate the transactional records of sales and marketing; a variety of other sources must be considered, such as end-user habits, real-time mobile and IoT feeds, and other diverse near-real-time unstructured sources. The traditional on-premises EDW is not agile or scalable enough to keep up with the ever-changing demands of today's next-generation data warehouse (DW) requirements. For IT agility and effective global consolidation of data sources, organizations must look to the cloud.

ESG research reveals that improving data analytics for real-time BI and customer insight is a top business priority followed closely by cost reduction, with nearly a third of organizations reporting that each of these initiatives will drive the most technology spending at their organization in 2019 (see Figure 1). In fact, 54% of organizations plan to increase spend in support of BI, analytics, and big data. Cloud-first strategies for analytics, DW, and BI are becoming prevalent. BI/analytics and database workloads are on the list of top workloads that organizations believe are most likely to move to the cloud. Forty-seven percent of organizations who leverage IaaS/PaaS services in the cloud do so for the purpose of running BI queries.[1]

Figure 1. Top Five Business Initiatives Driving Technology Spending in 2019



Source: Enterprise Strategy Group

While there are many cloud-based EDW solutions to choose from, the implementation of these solutions can be quite different. Choosing the right EDW solution requires understanding not only the underlying technology, but also the operational capabilities and pricing models.

[1] Source: ESG Mastery Survey Results, *2019 Technology Spending Intentions Survey*, March 2019.

The Solution: Google BigQuery Serverless Enterprise Data Warehouse

Google BigQuery is a cloud-based, fully managed, serverless enterprise data warehouse that supports analytics over petabyte-scale data. It delivers high-speed analysis of large data sets while reducing or eliminating investments in onsite infrastructure or database administrators. BigQuery scales its use of hardware up or down to maximize performance of each query, adding and removing compute and storage resources as required.

Google BigQuery, part of the Google Cloud Platform, is designed to streamline big data analysis and storage. Some of the specific advantages of Google BigQuery for businesses that work with big data include:

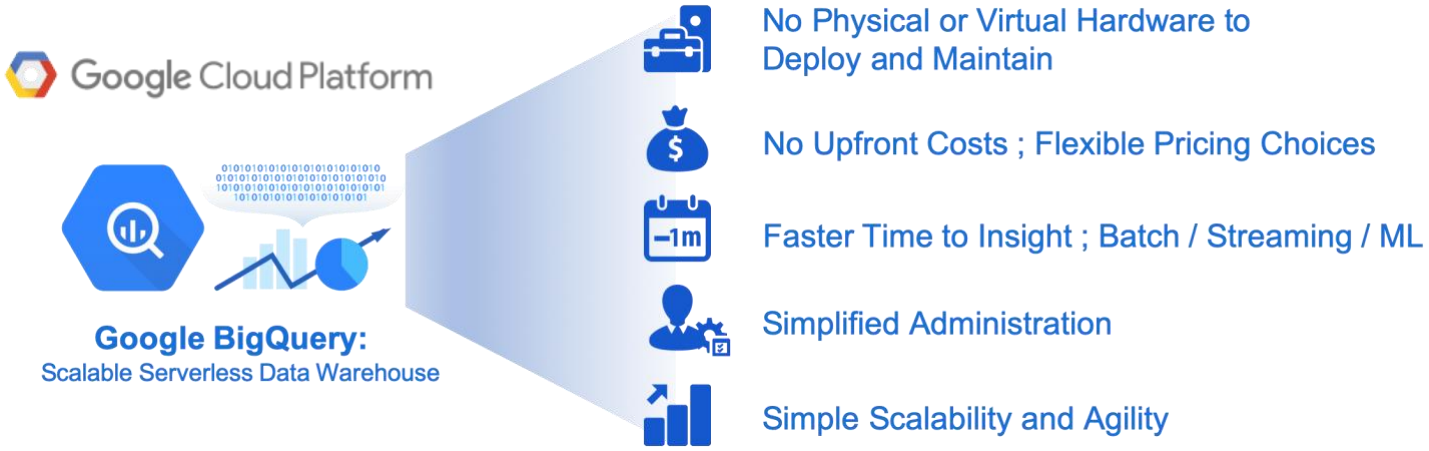
- **Time to value –** Get the data warehouse environment online quickly, easily, and without expert-level system and database administration skills by eliminating the infrastructure and reducing the management (known as "No Ops" or "Zero Ops").
- **Simplicity –** Complete all major tasks related to data warehouse analytics through an intuitive interface without the hassle of managing the infrastructure.
- **Scalability –** Scale up to petabytes or down to kilobytes depending on size, performance, and cost requirements.
- **Agility –** Act faster on new business opportunities, explore data sets to uncover new insight, leverage BQML, and innovate beyond the traditional expectations of a data warehouse.
- **Speed –** Ingest, query, and export PB-size data sets with impressive speeds using the Google Cloud Platform as the underlying cloud infrastructure.
- **Reliability –** Ensure always-on availability and constant uptime running on the Google Cloud Platform with geo-replication across Google data centers.
- **Security –** Protect and control access to encrypted projects and data sets through Google's cloud-wide identity and access management (IAM).
- **Cost optimization –** Predict costs with transparent and flexible flat rate and pay-as-you-go pricing options and contain costs through the use of project and user resource quotas.

Google BigQuery is self-scaling; it identifies resource requirements for each query to finish quickly and efficiently and provides those resources to meet the demand. Once the workload has completed, BigQuery reallocates those resources to other projects and other users. While both transferring data in and processing that data for results, BigQuery delivers tremendous speeds even at petabyte scale.

For enhanced data durability, BigQuery provides high availability and reliability through geographic replication that is completely transparent to its users, without the requirement to obtain the physical resources and space to house it all. BQ provides automatic backup and snapshots of your data sets, allowing for point in time restoration and ensuring that VPC service controls and data encryption (both in transit and at rest) are built-in to provide enhanced levels of security.

Ultimately, Google BigQuery enables organizations to address the cost and complexity challenges associated with building and maintaining a fast, scalable, and resilient big data infrastructure. By leveraging Google BigQuery's cloud-based approach, the time and cost traditionally dedicated to protecting data and guaranteeing uptime is nearly eliminated. With Google handling scalability, replication, protection, and recovery, organizations can focus more on gaining valuable insights, as opposed to infrastructure management.

Figure 2. Google BigQuery



Source: Enterprise Strategy Group

Google BigQuery versus Alternative Cloud-based EDW Services

ESG compared Google BigQuery's serverless EDW solution with some alternative EDW services offered by AWS, Microsoft, and Snowflake. While all of these offerings provide significant cost savings, reduction in complexity, and increased business agility when compared with an on-premises EDW solution, there are some significant differences between the offerings.

Amazon Redshift: Redshift is AWS' cloud-based and fully managed data warehouse service. Like an on-premises cluster, Redshift is based on the concept of nodes (virtual nodes) that must be deployed, configured, and managed. The management is greatly simplified since there is no hardware to physically administer and maintain. To scale the deployment, similar nodes of a fixed compute and storage capacity are added simultaneously, sometimes resulting in provisioning more compute or storage capabilities in order to meet the requirements of the other. Redshift clusters require planning and sizing, configuration and access through a "leader node," and scheduled updates several times per year. While AWS does offer many flexible pricing options, to achieve the greatest savings, organizations must pay for three years of service upfront, tying them in to both the vendor solution and the particular instance type that they have chosen.

Microsoft Azure SQL Data Warehouse: Azure SQL Data Warehouse is a cloud-based EDW solution that leverages a massively parallel processing (MPP) architecture to process Polybase T-SQL queries. Compute and storage (Azure Blob Storage) resources are separated and can be scaled independently. Azure SQL DW makes use of a control node that automatically directs parallel queries across all other compute nodes and moves data between those nodes as required. Compute resources are sold according to predefined service levels in Data Warehouse Units (DWUs) that provide a given set of compute resources for use in processing queries.

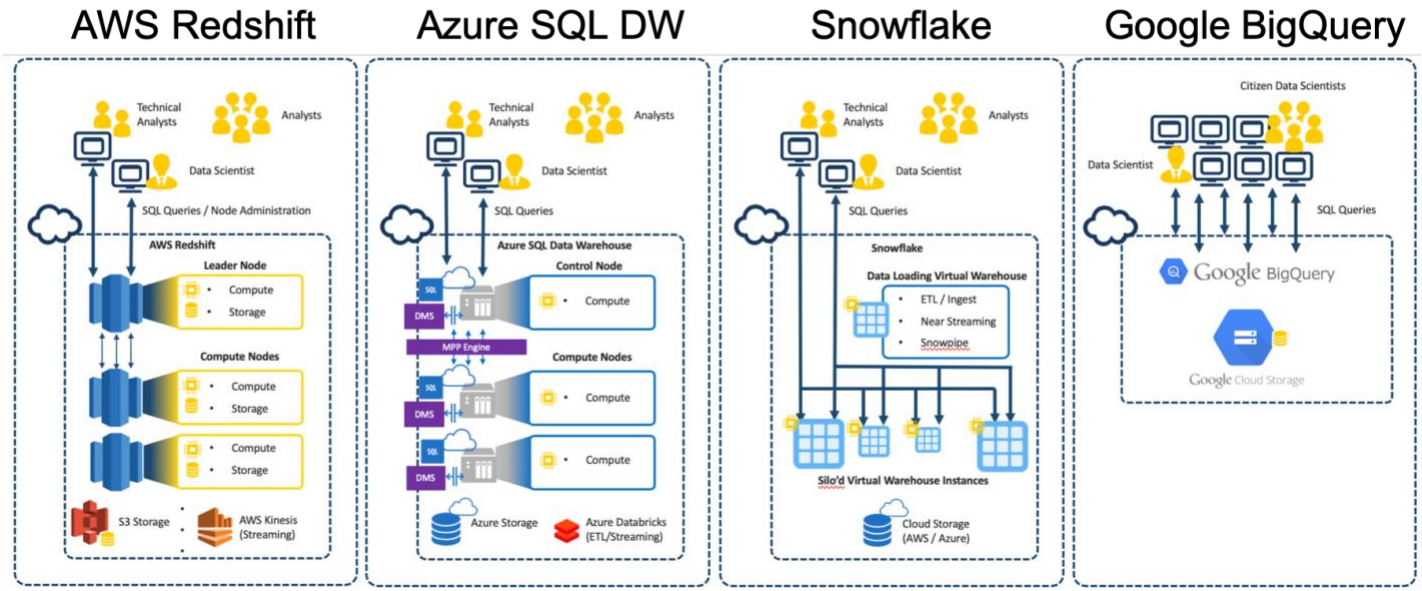
Snowflake: Snowflake is a managed data warehouse-as-a-service (DWaaS) that can be deployed on AWS or Azure infrastructure. Snowflake does not require management of physical or virtual hardware, installation of software, or maintenance. Snowflake also separates compute and storage resources and makes use of an MPP architecture behind the scenes. At the time of deployment, customers must select from preconfigured virtual data warehouses in various sizes (small, medium, large, x-large, etc.), which are priced based on "credit-hours" consumed. Per-credit cost varies based on region, cloud service, and support level. Virtual warehouses are usually siloed based on business unit and can be paused and resumed automatically to limit cost and can be resized when required. Snowflake also suggests making use of a dedicated virtual warehouse to handle data loading and ingest to avoid contention in resources required for queries.

Google BigQuery Serverless EDW: Google's BigQuery solution is completely serverless, self-scaling, self-maintaining, and self-tuning. There are no nodes to plan, configure, or scale. The complexity of sizing, managing, and maintaining the

physical infrastructure is handled behind the scenes by Google, removing the burden from the organization. End-users gain the benefit of all the auto-tuned and optimized resources working simultaneously. Customers can either pay by the total amount of data processed per month or opt to pay a flat rate fee based on number of "slots" (effectively the reserved quantity of parallel resources that are made available for running queries). While slots can be allocated to a particular department to process queries, unused slots may be allocated to other departments to handle bursts, optimizing slot utilization. Improvements to optimize queries are added monthly with a goal of shortening query execution time and minimizing the amount of data processed—thus minimizing on-demand costs to the end-user. BigQuery also supports native SQL-based machine learning (BQML) and GIS analysis, allowing all users to benefit from the power of advanced analytics without the need to integrate with other solutions.

Figure 3 depicts the four solutions compared in this analysis.

Figure 3. Functional Comparison of Cloud-based EDW Solutions



Source: Enterprise Strategy Group

ESG Economic Validation

ESG's Economic Validation process is a proven method for understanding, validating, quantifying, and modeling the economic value propositions of a product or solution. The process leverages ESG's core competencies in market and industry analysis, forward-looking research, and technical/economic validation. To validate the assumptions and costs included in the analysis, ESG conducted in-depth interviews with and reviewed case studies of organizations that had previously migrated their operations off of a legacy on-premises EDW solution and into BigQuery. This helped to better understand and quantify how the change to BigQuery impacted or affected their organizations. We used our findings to create a detailed economic model comparing the expected costs and benefits of BigQuery against the competitor's on-premises and cloud-based solutions.

Economic Value Overview – Google BigQuery versus Alternative Cloud-based EDW Solutions

ESG's economic analysis revealed that BigQuery can provide significant capital and operational savings and tangible benefits when compared with other cloud-based EDW solutions. ESG found that BigQuery provided customers with significant savings and benefits in these categories:

- **Elimination of upfront investment and planning –** BigQuery's serverless design is billed monthly with flexible on-demand or flat-rate pricing that eliminates the need to pay months or even years in advance to reduce cloud spend and eliminates the guesswork and planning required to size resources correctly. This gives the organization the greatest business agility when it comes to cloud services costs.

- **Reduction in operational expenses –** BigQuery eliminates the need to manage virtual EDW nodes as well as the need to monitor, troubleshoot, update, tune, and plan for growth. Google cloud storage is automatically optimized for cost, patching and maintenance is not required, and the support team is well trained and responsive. This leaves administrators more time to focus on other areas of the business.

- **Greater business agility and reduction in cost of daily administration –** BigQuery scales up or down as needed to meet the changing business demands, enabling organizations to quickly act on new opportunities without the need to plan configuration requirements, pause databases, or spin up dedicated warehouse instances for each organization. The solution also helps to eliminate or reduce the time spent on database administration, ETL management, and new schema modification. BigQuery also is the only solution that provides native AI/ML and supports native integration with many other cloud-based services.

Elimination of Upfront Investment and Planning Deployment Size

Migrating from an on-premises EDW to any cloud-based EDW solution eliminates the need to make a large upfront investment in a physical infrastructure and helps to avoid overprovisioning physical compute and storage hardware for growth and spikes in demand. However, not all cloud-based EDW solutions completely avoid the need to make upfront investment or the need to plan the deployment size or overprovision to handle spikes in workload.

- **Elimination of upfront investment** – BigQuery and Snowflake require no upfront investment to achieve the greatest level of savings for compute resources. Snowflake does provide lower storage costs to those that reserve capacity, but this does not require an upfront investment (the customer is, however, required to pay for unused storage). Both AWS Redshift and Azure SQL DW require annual contracts and upfront investment to achieve the greatest savings. This money must be paid one or three years in advance, even though the value of the service is not realized until months or years later. Organizations should be sure to take the additional cost of capital into consideration when comparing upfront on on-demand costs.

- **No planning of deployment size –** With BigQuery, customers do not have to plan and size the configuration to meet their needs. BigQuery is completely serverless and customers do not have to plan or adjust the supporting infrastructure. With AWS Redshift, the size and quantity of the instances required to handle ingest and workload must be predicted. With Azure SQL DW, users must choose the cDWU rating that best meets their needs. Snowflake users must determine the mix of storage and the size of the virtual data warehouse that best suits the needs of each business unit as well as for a dedicated warehouse used for loading data.

- **No need to grow deployment –** BigQuery does not require monitoring of capacity utilization or growing the environment when limits are reached. Scaling up resources for Azure SQL DW and Snowflake are relatively easy but

require some degree of planning and administrative activity. AWS Redshift deployments can be scaled up, but this requires not only management overhead, but also some temporary downtime of the cluster as instances are added.

- No need to overprovision – The BigQuery infrastructure utilizes all resources available to process queries and does not have to be overprovisioned to handle unexpected spikes in workload. AWS Redshift, Azure SQL DW, and Snowflake Virtual Warehouses define a fixed set of resources that are available for queries to run against and must be overprovisioned to handle ingest, reclustering, and vacuum processes. Under that system, a configuration that is too small negatively impacts queries, leading organizations to size and pay for deployments sized to handle the worst case scenario.

Table 1. Comparing Cloud-based EDW Solutions: Upfront Investment, Planning, and Agility

	Google BigQuery	AWS Redshift	Azure SQL DW	Snowflake
Upfront Investment	None.	One-year or three-year contracts with upfront investments required to receive competitive rates.	One-year or three-year reservations required to receive discounts.	None.
Sizing and Planning	No planning required.	Must predict correct instance sizes and reserved instances limit agility.	Predict cDWU required. Can scale up/down easily on demand, but reservation limits agility.	Predict virtual warehouse size required for each business unit.
Agility / Growth	Increase slots if needed.	Resizing cluster size requires cluster downtime.	Add cDWU as needed.	Scale by adding new virtual warehouses or growing the size of individual virtual warehouse.
Overprovisioning for Capacity Growth or Performance Spikes	None.	Must overprovision enough instances to handle worst case scenario or increase operational overhead and downtime of scaling up and down nodes as needed.	Must overprovision enough instances to handle worst case scenario or increase operational overhead of choosing new cDWU deployment size as needed.	Must overprovision enough instances to handle worst case scenario or increase operational overhead of scaling up and down warehouses as needed.

Source: Enterprise Strategy Group

Reduction in Operational Expense

Eliminating the need to manage and maintain hardware on-premises brings about substantial operational savings for the organizations that migrate to cloud-based services. The completely serverless nature of BigQuery brings about additional operational advantages, even when compared with other cloud-based EDW services aimed at reducing these costs.

- **Eliminates node management –** BigQuery is completely serverless and stateless. AWS Redshift requires virtual node management for tasks such as configuring network access and VPNs, allocating resources, updating software, maintaining access to resources, and growing and shrinking the cluster. Azure SQL DW requires maintenance, operation of a control node, and manual scaling. Snowflake is a little more hands off than the others, but still requires manual scaling and management of multiple virtual warehouses.

- **Reduced storage management –** BigQuery is the only cloud-based EDW solution of the four that offers self-optimizing storage. Data older than 90 days is automatically moved to cost-effective storage to automatically reduce costs.

- **Zero compute maintenance –** BigQuery and Snowflake are maintained non-disruptively to the user in the background. Snowflake nodes keep state, and bringing them down can impact performance. Azure SQL DW and AWS Redshift require that manual maintenance windows be scheduled, and updates can result in downtime.

- **Support –** While not a clear operational advantage for any cloud-based solution, it should be noted that Google, AWS, and Microsoft have large established support operations and professional services organizations, while Snowflake is trying to scale to meet current and future support demand.

Table 2. Comparing Cloud-based EDW Solutions: Operational Expenses

	Google BigQuery	AWS Redshift	Azure SQL DW	Snowflake
Virtual Node Management	None; completely serverless.	Management node required with manual configuring and scaling of nodes.	Control node required with simple scaling of data warehouse compute units.	Serverless, but may have to manage siloed data warehouses, and a dedicated warehouse to load data is suggested.
Storage Management	Self-optimized storage automatically moves data to cost-effective storage.	Manual storage migration and aging.	Single storage tier.	Must manage data movement and decide between expensive fast storage and economical capacity storage.
EDW / Node Maintenance	Managed by Google in the background with no downtime.	Manual updates of nodes during scheduled downtime.	Updates of nodes during scheduled downtime.	Managed by Snowflake in the background with no downtime.
Enterprise-level Support	Percentage of cloud spend with minimum spend (lesser support levels available).	Percentage of cloud spend with minimum spend (lesser support levels available).	Requires customized quote for premier support (lesser support levels available).	Priced into hourly compute credit cost (lesser support levels available).

Source: Enterprise Strategy Group

Improved Business Agility and Reduction in the Cost of Daily EDW Administration

Daily EDW administration often consists of tasks such as managing the ETL process between systems; configuring, managing, and monitoring the platform and database; troubleshooting; controling access; maintaining security; developing new services; collaborating with business analysts; providing data for reports and dashboards; and integrating applications and other cloud services. ESG found that when compared with an on-premises solution, BigQuery and the other cloud-based solutions can eliminate some tasks altogether, and make other tasks simpler and quicker through automation and integration. BiqQuery provides additional management simplicity, even when compared to the other cloud-based services:

- Eliminated need to monitor and resize virtual resources – BigQuery administrators do not need to spend time monitoring or scaling the virtual resources of the system (nodes, compute units, or virtual warehouses). Since BigQuery concurrency limits are soft, they can easily be increased when needed. For those that do wish to monitor the BiqQuery environment, many functions allow for insight into query performance, trends, billing, and system audits with the ability to create automated alerts based on thresholds.

- Improved business agility – Responsive scalability means that queries can be completed sooner, informed decisions can be made faster, and resources to satisfy time-sensitive business opportunities are available at all times. Resources do not have to be allocated or expanded to meet the needs of new opportunities. BQ flat rate customers can share idle resources between slot pools to handle bursts. BigQuery customers reported that they were much more agile and able to act on opportunities much faster with no impact to the business.

- Simplified ETL and management of data sources – The customers whom ESG spoke with reported that flexible options allow them to stage data either on-premises or in the Google Cloud, and store data in a variety of options within the Google Cloud (Cloud Spanner, BigTable, Cloud Storage, etc.) or from external data sources. These flexible options allow for greater possibilities and capabilities through the use of other Google products. Some alternative options exist on the AWS and Azure ecosystems, but Snowflake has limited options.

- Simplified data ingest – Data ingest is one of BigQuery's strengths when compared with the alternative cloud-based EDW solutions. ESG reviewed the results of credible third-party testing that showed BigQuery scored 1.9X to 2.4X higher than the alternative solutions at data ingest. Data ingest is free and consumes zero user resources on BigQuery. AWS Redshift, Azure SQL DW, and Snowflake all require far more complex configurations and tasks to manage ingest operations, and ingest operations leverage user compute resources, essentially requiring overprovisioning of compute resources to meet workload demand. The Snowflake service suggests deploying a dedicated virtual warehouse for data loading and ingest, adding both cost and management complexity.

- Simplified legacy EDW-related tasks – BigQuery customers reported saving time by eliminating backups and system maintenance tasks, simplifying processes such as onboarding, managing workload priorities, and maintaining partitions.

- Greater flexibility of SQL queries and schemas – BigQuery, AWS Redshift, and Azure SQL DW often do not require any schema change from most existing DW solutions, but BigQuery also provides for some enhanced functions that can help optimize schemas and make queries faster and more powerful. Snowflake also offers optimizations but leverages proprietary code that requires learning and experience.

- Simplified access control and security – Google Cloud Deployment Manager can help automate creation of IAM custom roles, which provide simplified project management and access control via the flexibility to quickly manage access to resources by functional role, organization, data lifecycle stage, or project. BiqQuery eliminates the need to manage individual users and organize and grant permissions, and has security and encryption features built in.

Table 3. Comparing Cloud-based EDW Solutions: Improved Agility/Simplified EDW Management

	Google BigQuery	AWS Redshift	Azure SQL DW	Snowflake
Plan and Size Virtual Resource Requirements for New Workloads	No additional sizing required.	Size AWS instances, considering compute, memory, and local storage.	Size additional DWU requirement.	Size new siloed virtual warehouse or scale existing warehouse.
Provision Cluster / Virtual Warehouses / Networking	No provisioning is required.	Provision VPCs, network, user access, and storage resources on new instances. Downtime required.	Configure instances, access, and firewalls	Simple provisioning.
Update Schema / Scripts / Queries[2]	Few changes required for many standard solutions.	No changes required for most standard solutions.	SQL Server customers will require no changes.	Few changes required for many standard solutions.
Manage / Monitor / Scale Virtual Resources	None required.	Monitor cluster for performance and scale up/back as needed. Reserved instances cannot be scaled back.	Monitor cluster for performance and scale up/back or pause/resume as needed. Reservations cannot be scaled back.	Monitor cluster for performance and scale up/back/out. Reliance on caching can result in unpredictable performance.
Manage User Access / Security	Simplified IAM controls; role-based access with automatic patching.	Manage users and security for each cluster. Patching during scheduled windows.	Manage firewalls and AD authentications. Patching during scheduled windows.	Simplified role-based and object level permissions with automatic patching.
Manage Data Loading / Ingest / Streaming Data	Flexible data ingest, transformation, and streaming options. Highest performance network.	Complex data loading and streaming requires Kenesis service. Ingest consumes allocated compute resources.	Complex data loading and streaming requires Azure DataBricks cluster. Ingest consumes allocated compute resources.	Separate warehouse is suggested to be running for data loading purposes. Snowpipe service is only near real time.
Integrate with Other Cloud-based Tools and Services	Designed for integration with many Google services including DataFlow, DataProc, CloudDB, PubSub, and sheets, requiring minimal management overhead.	Manual integration with AWS services; some management and configuration overhead.	Manual integration with Azure services that requires management and complexity overhead. Some redundancy of resources.	Bolt-on integration with services from other vendors requires management and operational overhead.
Integrate with AI / ML Workloads	Designed to be used in conjunction with AI/ML workloads with SQL-based ML.	AI/ML requires integration with other solutions and transformation of the data.	AI/ML requires integration with other solutions and transformation of the data.	No AI/ML workload support; must use other solutions.

Source: Enterprise Strategy Group

[2] Some changes to existing on-premises schema, scripts, or queries may be required depending on solution and support.

Migration Considerations

Migrating on-premises EDW workloads to a cloud-based EDW solution or service provides significant advantages and opportunities for cost savings, but this cannot be realistically achieved overnight. Organizations should analyze what tasks are involved and devise a solid plan for deciding when to move their EDW to the cloud.

Migration strategy for EDW: Organizations should consider whether they wish to migrate data all at once and flip the switch on existing operations or migrate operations slowly over a period of years while phasing out the on-premises EDW. Organizations should consider the costs involved to migrate in terms of time, physical transfer of data, professional services, potential downtime, etc.

- Development and testing costs – Organizations should consider the costs involved in retooling applications, transforming data into new schemas, rewriting optimized queries, testing, validating, troubleshooting, creating custom applications, retraining developers, etc.
- Process redesign – Change presents the opportunity to improve many processes and systems. Organizations should consider the cost in terms of time and money to make changes to the ETL process, including streaming capabilities, integrating with other cloud products, and educating business teams and IT resources on the new process and tools.
- Server and software costs – Organizations should consider the costs associated with supplemental resources such as on-premises or cloud-based staging servers or software and SaaS licenses, without forgetting to balance these new costs against any potential savings gained by migrating to the cloud.

To help ensure a successful migration, Google Cloud provides organizations with a prescribed migration framework with resources, funding, and a proven methodology.[3]

ESG Three-year Modeled Scenario

ESG leveraged the information collected through vendor-provided material, public and industry knowledge of economics and technologies, and the results of customer interviews to create a three-year TCO/ROI model that compares the costs and benefits of satisfying a modeled organization's EDW requirements with Google BigQuery versus three other cloud-based EDW solutions. ESG's interviews with customers who have experience with multiple cloud EDW solutions and publicly available pricing information and sizing guidelines, combined with our experience and expertise in economic modeling and technical validation, helped to form the basis for the assumptions used in our modeled scenario.

To reflect the size and scale of the customers interviewed, ESG considered a scenario of a typical large organization (over 10,000 employees) with requirements to store and analyze approximately 200 TB of data generated by several business units. Of this data, 110 TB was assumed to be a fixed size data set of historical and updated tables that did not expire, and ESG assumed that 250 GB of data was ingested daily with a retention period of 365 days. ESG assumed that the BigQuery deployment was priced at an annual flat rate capacity of 2,000 slots. The other solutions were sized using conservative estimates of relationships between number of BigQuery fixed slots and ECU (AWS Redshift), DBU (Azure SQL DW), and Snowflake deployment size. These relationships were obtained in good faith by analyzing information collected in the field from head-to-head POCs, solution replacements, and published data. In practice, the competitive solutions may require even more resources be deployed to deal with the additional resource overhead required for data ingest, reclustering, and vacuum processes, which are free of charge and have no negative impact to performance on BigQuery.

[3] https://cloud.google.com/blog/products/data-analytics/migrating-teradata-and-other-data-warehouses-to-bigquery

The configuration assumptions for each solution are shown in Table 4. To reflect the benefit of storing compressed data versus uncompressed data, Google BigQuery required the full 200 TB of data be stored on Google Cloud Storage, while all of the other solutions required only half of this capacity (~100 TB) based on the assumption of 2:1 compressibility of the data. For AWS Redshift, the latest generation dense compute nodes were used and only the remainder of the required capacity that was not able to fit on the included SSD space was priced as S3 storage.

Table 4. Configuration Assumptions Used in ESG Three-year Modeled Scenario

	Google BigQuery	AWS Redshift	Azure SQL DW	Snowflake
Compute / Service Pricing Option	Flat Rate Pricing (Annual Agreement) (paid monthly)	One-year Reserved Instance Pricing (42% Savings) (paid upfront)	One-year Reserved (37% Savings) (paid upfront)	Enterprise+ (AWS US East) ($4.00/Credit)
Compute / Service Configuration	2,000 Fixed Slots	16 x dc2.8xlarge instances	DW6000c	44 Credits (assumed average utilization of 12hrs/day)
Annual Cost of Capital for Upfront Spend	N/A	8%	8%	N/A
Cloud Storage	199.1TB Google Cloud Storage (auto optimized for cost savings)	19.6 TB of usable storage included with instances plus 79.9 TB of S3 storage	99.56 TB of Azure Blob Storage	99.56 TB of Capacity Storage (paid upfront for 43% savings)
Streaming Service / Data Loading	Streaming Inserts	AWS Kinesis	Azure Databricks (Standard tier, 6 DBU)	Large Data Warehouse (8 Credits x 8hrs/day)
Support Level	GCP Business-critical	AWS Enterprise Support	Azure Premier Support	Enterprise+ (Support included in credits)

Source: Enterprise Strategy Group

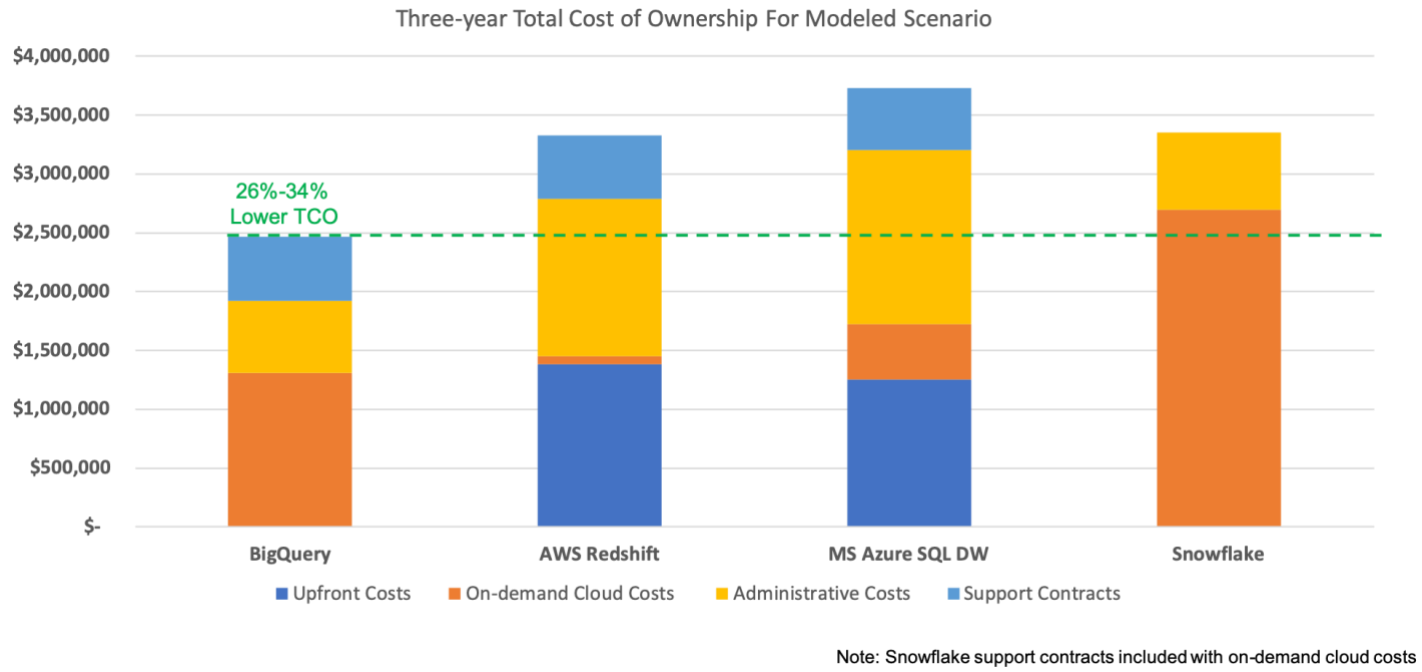
Because this was not designed as a study of moving on-premises workloads to the cloud, ESG did not estimate the time and cost of initial solution migration and complexity for each vendor in this analysis.[4] We assumed it to be equal for all solutions. ESG also assumed that the performance supplied by these configurations would satisfy the SLAs of the business, and no direct performance advantage was given to one solution over the other. In practice, these factors can vary widely based on your environment and should be carefully considered by your organization. BigQuery may in fact have

[4] ESG studied the cost of moving on premise workloads to the cloud in *The Economic Advantages of Migrating Enterprise Data Warehouse Workloads to Google BigQuery, March 2019*

advantages in both of these aspects, but we conservatively did not assign any cost advantage to any solution based on time and complexity to migrate or improved/more predictable performance.

Based on our conservative assumptions, ESG found that the Google BigQuery solution provided a 26-34% lower total cost of ownership than AWS Redshift (26% lower), Microsoft Azure SQL Data Warehouse (34% lower), and Snowflake (27% lower) over the modeled three-year period. Figure 4 shows the estimated three-year total cost of ownership for each of the four cloud-based EDW solutions to satisfy the needs of the modeled scenario.

Figure 4. Estimated Three-year Cloud-based Data Warehouse Solution Total Cost of Ownership (TCO)



Note: Snowflake support contracts included with on-demand cloud costs

Source: Enterprise Strategy Group

Upfront Cloud Spend (Prepaid Capital Investment)

When estimating cloud spend, we conservatively compared Google's flat rate pricing with a cost-effective solution for each of the three competing solutions over the three-year period. For AWS Redshift and Azure SQL DW, this involved signing agreements and paying upfront one year in advance to achieve the reported level of savings (37-642% savings) for compute resources. For payments made one year in advance, we calculated the financial impact of paying for the service before the benefit was received by calculating a conservative 8% cost of capital to any portion of the remaining payment in which the benefit has not yet been realized (similar to a diminishing interest payment calculation). As a result, the overall cumulative cloud spend for AWS and Azure SQL DW was slightly lower than for BigQuery. It should be noted, however, that by reserving compute in advance, the solution becomes far less flexible, making it more difficult to take advantage of improved technologies or scale back the deployment to lower costs. Discounts for AWS and Azure can be increased further by reserving resources for up to three years in advance. However, this further reduces the flexibility of the solution. It should also be noted that ESG's models predict that when selecting on-demand pricing for all solutions, BigQuery's expected three-year TCO was estimated to be 60% to 69% lower than the other cloud-based solutions.[5] Google BigQuery and Snowflake required no upfront payment.

[5] To estimate BigQuery on-demand costs (billed per TB queried), our models assumed each ingested GB of data was queried 100 times per month

On-demand Cloud Costs

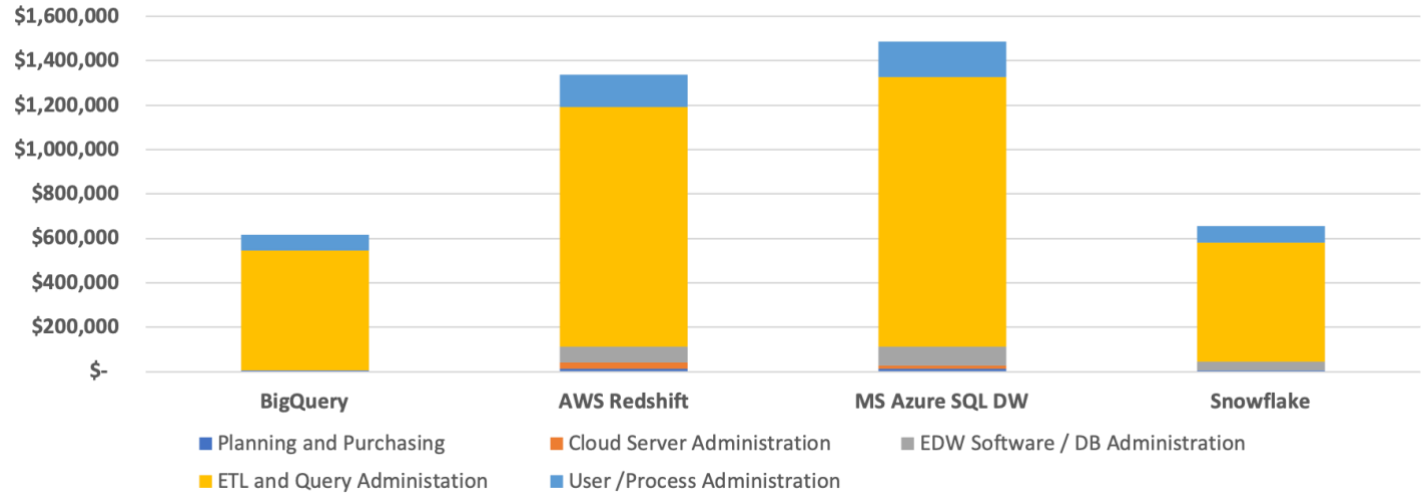
Monthly cloud spend consisted of the estimated bill that would be received by the organization each month for compute, storage, and/or services. The compute cost of AWS Redshift was all paid upfront annually, and the instances selected each provided roughly 20% of the total storage capacity required. Monthly cloud spend for AWS Redshift consisted only of 79.9 TB of S3 storage capacity and the estimated cost of Kinesis streaming service to handle 760 GB of streaming data. The cost of Azure SQL DW compute was also paid upfront annually. Azure SQL DW monthly cloud spend consisted of 99.6 TB of Azure Blob storage and a modest Azure Databricks setup (6DBU) to handle ETL and steaming operations. Snowflake monthly spend consisted of 528 daily credits (44 credits x 12 hrs/day) for virtual warehouses, an additional 64 daily credits (8 credits x 8 hrs) for the data loading virtual warehouse and Snowpipe service, and 99.6 TB of capacity storage. Snowflake monthly cloud spend also includes the cost of support built in, so they are reported here rather than in support/maintenance. Google BigQuery costs included 2,000 fixed rate slots, optimized cloud storage costs (data older than 90 days moved to more cost-effective storage), and streaming inserts.

Administrative Costs

Administrative costs for each of the solutions were established on good faith estimates based on expert and end-user opinions and case study analysis. ESG modeled the expected one-time or weekly hours that an employee would spend on planning and purchasing, administration of virtual cloud servers or nodes, EDW software and DB administration, ETL and query administration, and user/process administration. All of the cloud-based solutions provided significant administrative savings when compared with ESG's on-premises models. Planning and purchasing for all four solutions is greatly simplified in comparison to planning on-premises deployments, but AWS Redshift and Azure SQL DW each would require some time spent on sizing and analyzing the solution before making a three-year upfront investment, while nonideal sizing of the BigQuery or Snowflake environment could be easily remedied at any time.

Once deployed, the Redshift and Azure SQL DW solutions would require monitoring, configuration, and tuning of virtualized resources (instances or nodes), while both the BigQuery and Snowflake solutions are fully managed behind the scenes. BigQuery was the only solution that did not require tuning and configuration of the EDW software and database. BigQuery and Snowflake were reportedly easier to manage from an ETL, query administration, and user/process administration perspective, and this provided the majority of the administrative savings in our analysis. BigQuery simplified ingest, and the ability to support native AI/ML and integrate natively with other tools provided additional administrative savings. Figure 5 compares the modeled three-year administrative costs for each of the four solutions.

Figure 5. Estimated Three-year Administrative Spend for ESG's Modeled Scenario



Source: Enterprise Strategy Group 2019

Support Contracts

Enterprise-level support contracts for BigQuery and Redshift are similarly priced and based on monthly cloud spend with a minimum of $15K/month. Azure Premier support pricing is not published and requires a customized quote, but ESG estimated it (with low confidence), based on available information, at $175K/annually. While it does not have a major effect on the bottom line of the analysis, Azure support costs may be lower than reported in our analysis. Snowflake support is built into the per-hour credit cost, and thus could not be broken out and reported separately. ESG's model included Snowflake support in the on-demand cloud cost category.

The Bigger Truth

It should be a top priority of every organization to produce quality business intelligence and actionable insight as quickly and cost-effectively as possible. Cloud-based data warehouse services offer IT organizations significant advantages in agility and availability, while greatly reducing time to insight. Whether your organization is looking to move your entire on-premises data warehouse operation to the cloud, make a slow transition, or augment existing operations with cloud-based EDW services to handle new opportunities, it is important to understand the differences between offerings and choose the solution that provides you the greatest agility, flexibility, and interoperability with other services at the lowest cost.

ESG compared the cost and capabilities of four leading cloud-based EDW solutions: Google BigQuery, AWS Redshift, Azure SQL Data Warehouse, and Snowflake. ESG leveraged interviews with end-users who had experience with multiple solutions as well as case studies to form conservative assumptions used to model the total cost of ownership (TCO) to satisfy the needs of a modeled organization's EDW requirements over a three-year period. Our model predicted that BigQuery can provide a 26-34% lower TCO in a solution that provides simpler operation and greater agility and scalability, with native capabilities and interoperability with other important cloud-based services.

Google BigQuery is designed to remove all of the physical and logical burden of managing, monitoring, maintaining, and securing EDW infrastructure, allowing organizational resources to focus on obtaining intelligence rather than the process of generating it. BigQuery is the only fully managed solution (no physical *or* virtual nodes to manage and maintain) with self-optimized storage and native support for AI/ML workloads. BigQuery provides savings over the other cloud-based EDW solutions by eliminating the needs to size the deployment, manage the environment, and plan for growth of virtual resources, and by providing simpler operation and daily administration. In addition, BigQuery requires no upfront investment and can easily scale up or down to meet the ever-changing needs of the business.

Today's EDW solutions must act as a global repository of information, provide the agility to scale up or down with demand, and seamlessly integrate with other analytics and operational functions. But simply comparing public price sheets is not enough—you must consider the cost of people and process as well. ESG recommends every organization perform its own analysis to compare solutions, and we hope that this report helps to identify the areas that should be considered when comparing cloud-based EDW services. We also strongly recommend that you consider Google BigQuery in conjunction with other cloud-based services from Google to power the platform that provides next-generation business insight across your entire organization.

Enterprise Strategy Group is an IT analyst, research, validation, and strategy firm that provides market intelligence and actionable insight to the global IT community.

www.esg-global.com

contact@esg-global.com

P. 508.482.0188