ESG Economic Value Validation

The Economic Advantages of Migrating Enterprise Data Warehouse Workloads to Google BigQuery

By Aviv Kaufmann, Senior Validation Analyst; and Alex Arcilla, Validation Analyst
March 2019

Executive Summary

For generations, organizations leveraged Enterprise Data Warehouse (EDW) systems to consolidate, store, analyze, and report business intelligence. A vendor who had launched the first parallel data warehousing system and the first DW system over 1TB in scale has had success delivering the hardware, software, and services that enabled large-scale on-premises data warehouse solutions for decades. These systems, though extremely costly to purchase and maintain, provided a business advantage to those organizations that could afford it. Today, the exponentially increased variety, volume, and velocity of data call for a data warehouse solution that is more agile, global, and cost-effective. Those who have previously relied on this vendor's solutions are faced with a question: continue to make investments in rigid, on-premises solutions, or face the one-time cost of migrating to an agile, cloud-based EDW solution.

ESG created a three-year total-cost-of-ownership (TCO) model that compared the expected costs and benefits of upgrading an on-premises EDW solution from the leading vendor mentioned above, migrating to a cloud-based solution provided by this vendor on AWS, or redesigning and migrating the EDW function to Google BigQuery. The costs and assumptions used in this modeled scenario were validated through in-depth interviews with organizations that had previously migrated their operations off of the legacy on-premises EDW solution and into BigQuery. ESG found that our modeled organization can reduce their overall three-year costs by 52% when compared to the on-premises solution and by 41% when compared to the solution on AWS. The elimination of the hardware investment and related operations and maintenance costs contributed largely to the decrease. Moreover, ESG also saw that the underlying architecture of Google BigQuery, decoupling processing capability and storage capacity, also contributed to lowering overall expenses.

Google BigQuery

41-52%
Reduction in TCO when migrating EDW deployments

The Challenges with Legacy On-premises EDW Solutions

The on-premises enterprise data warehouse (EDW) has been the backbone of many of the 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. One vendor emerged as a leader in providing the hardware and software technologies and accompanying services to successfully deliver some of the largest and most scalable on-premises EDW solutions. These solutions were perfect for consolidating the data provided by legacy operational systems. Significant planning was required to design, implement, manage, and maintain the hardware and rigid structure of the extract, transfer, and load (ETL) functions, but changes in sources of data were not common, and the solutions performed great for the function they were designed for.

Today, organizations are collecting data form 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, but also a variety of other sources from end-user habits, real-time mobile and IoT feeds, and a diversity of other near-real time unstructured sources. The traditional on-premises EDW is simply 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.

The large, expensive EDW that once provided a competitive advantage has become a business obstruction. The largest organizations understand that something must be done. Modernization of data analytics and business intelligence (BI) systems are a top priority. This involves designing BI systems that are flexible enough to quickly adapt to new sources of data, quickly scale up or down, and provide intelligence that is as close to real-time as possible. ESG research reveals that improving data analytics for real-time BI and customer insight is a top priority, with 31% of organizations reporting that it is one of the initiatives that will drive the most of their technology spending in the next 12 months. In fact, 54% of organizations plan to increase spend in support of BI, analytics, and big data.[1]

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.[2] Even the leading at-scale legacy EDW vendor described above has realized that, in order to continue to deliver modern, agile, and scalable solutions, they must focus on their software and services and partner with cloud service providers rather than continue to rely on on-premises solutions.

Customers that have come to rely on this vendor's on-premises DW solution face a choice: Update their current on-premises DW solution, move this solution to run on infrastructure-as-a- service (IaaS), or consider a modern, next-generation solution for DW.

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.

---

[1] Source: ESG Research Report, _2019 Technology Spending Intentions Survey_, February 2019.

[2] ibid

Google BigQuery, part of the Google Cloud Platform, is designed to streamline big data analysis and storage, while removing the overhead and complexity of maintaining onsite hardware and administration resources. Some of the specific advantages of Google BigQuery for businesses that work with big data include:

- **Time to value** – Users can get their data warehouse environment online quickly and easily, without requiring 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 your size, performance, and cost requirements.

- **Speed** – Ingest, query, and export PB-sized 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 flat rate and/or pay-as-you-go pricing 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 scales. 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.

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.

Google BigQuery versus Legacy EDW Provider

ESG compared Google BigQuery's serverless EDW solution to that of the leading legacy EDW provider discussed in the executive summary and *Challenges* section of this paper. Currently, this competitor offers both on-premises and cloud-based versions of its flagship solution.

Legacy EDW On-premises: The competitor's on-premises offering is comprised of commodity servers and all-flash storage, powered by its proprietary EDW software. Organizations must plan for a very large capital expense (millions to tens of millions of $USD) to cover the up-front costs in both hardware and software for the initial deployment, as well as any upgrades required to satisfy the unanticipated growth in storage or compute requirements. Although it varies, a typical investment in this on-premises EDW solution is sized to handle the expected requirements over a three-year period. As with other hardware-based solutions, organizations must plan, deploy, maintain, and configure the physical hardware and software required to store the data and power the queries. They also must spend considerable efforts configuring, administering, and optimizing this solution as data volumes and users scale over the life of the deployment.

Legacy EDW on Amazon Web Services (AWS) IaaS: The cloud version is deployed as virtualized instances of the vendor's EDW software running on AWS IaaS and sold in the AWS marketplace. AWS Elastic Compute Cloud (EC2) instances and Elastic Block Storage (EBS) are deployed and managed in the same manner as physical nodes are in the on-premises solution. While there is no cost of installing, deploying, managing, and maintaining physical hardware, the logical node-based system must be planned, deployed, maintained, upgraded, and operated by an administrator in the same manner as the on-premises solution. This solution is not self-scaling, nor self-tuning. To scale this cloud solution, organizations must add additional instances. However, incorrectly sizing the requirements of the system, or handling bursty or seasonal behavior may result in provisioning more storage and/or compute capacity than what is typically needed at steady state, resulting in significant overprovisioning, higher costs, and complexity.

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 number of parallel queries) made available. 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 being 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.

Figure 1 depicts the three solutions compared in this analysis and how they each implement the stages of the DW process.

Figure 1. Comparison of Legacy EDW Solution (On-premises, In the Cloud) and Google BigQuery
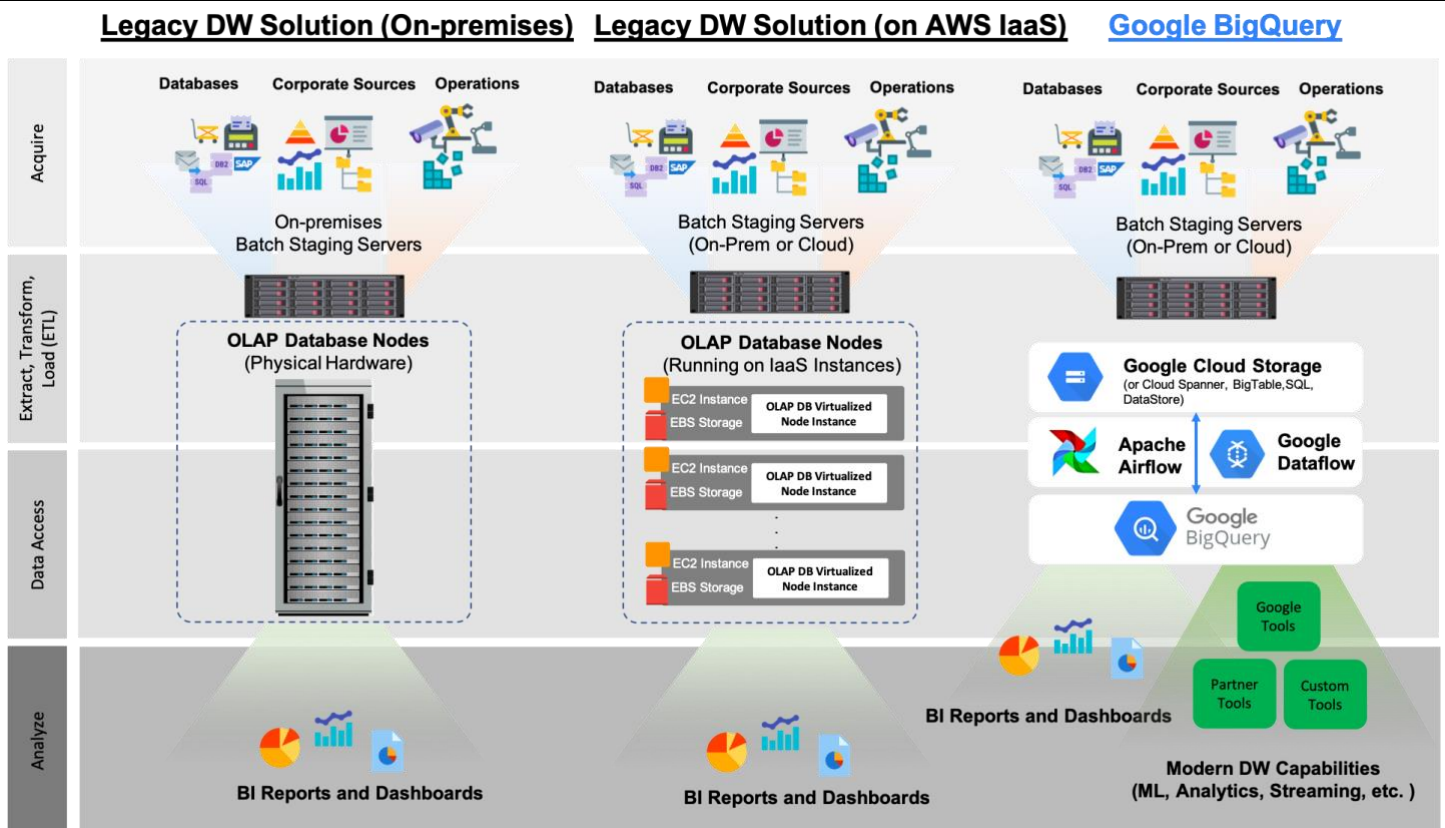
Source: Enterprise Strategy Group

ESG Economic Validation Process

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 the legacy on-premises EDW solution and into BigQuery. This helps to better understand and quantify how the change to BigQuery had 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 Leading Legacy EDW Solution

ESG's economic analysis revealed that BigQuery can provide significant capital and operational savings and tangible benefits when compared with a leading legacy EDW solution deployed either on-premises or on cloud infrastructure. ESG found that BigQuery provided customers with significant savings and benefits in these categories:

- **Elimination of up-front investment** – BigQuery's serverless design is billed monthly and eliminates the need to make up-front investments in hardware or sign annual contracts to reduce cloud-spend.

- **Elimination of on-premises related operational expenses** – BigQuery eliminates the need to spend on power/cooling/floorspace, maintenance, license management, and hardware and software upgrades.

- **Reduction in cost of daily administration** – BigQuery eliminates the need to manage physical or virtual EDW nodes, as well as the need to monitor, troubleshoot, update, tune, and plan for growth. The solution also helps to eliminate or reduce the time spent on database administration, ETL management, and new schema modification.

- **Improved agility, scalability, and availability** – BigQuery is able to scale up or down as needed to meet the changing business demands, enabling organizations to quickly act on new opportunities, provide faster time to insight, and perform queries that are not easily accomplished with legacy EDW solutions.

Elimination of Up-front Investment

ESG found that customers who had migrated their EDW to BigQuery avoided significant up-front capital investment in new hardware when compared to an on-premises EDW solution. This is perhaps the most obvious and significant savings that organizations will encounter. Deploying or upgrading a legacy EDW solution at scale represents an up-front investment of millions to tens of millions of US dollars. Some estimates of the legacy EDW provider solutions were as high as $15,000/TB. An investment of this magnitude pressures supporting teams to size the deployment correctly and provide metrics to justify the investment on a regular basis. One customer reported:

> *"Legacy data warehouses are large investments; a lot of people are very dependent on this. [This is] extremely visible to the organization."*

- **Elimination of EDW hardware spend altogether** – ESG found that customers that migrated to BigQuery eliminated their HW spend for an EDW solution altogether. There may be on-premises hardware required for staging, etc., but ESG's model assumes these will be similar regardless of EDW used.

"...We were spending a lot of time on capacity management, and every time it came to refresh the hardware - there was a large cost outlay that we had to work through."

- **Elimination of the need to overbuy EDW capacity and capability –** Customers confirmed that BiqQuery eliminated the need to project, procure, deploy, and maintain excess capacity to handle spikes, seasonal requirements, and growth. The organizations no longer had to spend on buying technology that may become outdated in two to three years. In fact, one organization stated that the need to begin the painful process of having to plan for growth in capabilities is what ultimately led them to consider migrating away from the legacy EDW solution.

	Legacy EDW On-premises	Legacy EDW on AWS	Google BigQuery
Up-front Investment	Millions to tens of millions of $USD.	Option to lower effective rate through annual contract. Pay all instance costs for the year up-front.	None.
Upgrades / Growth	Substantial capital investment to grow by adding new nodes.	Add instances, storage, and EDW software as new nodes are required.	Add slots or storage on demand.
Overprovisioning for capacity growth or performance spikes	Must purchase the hardware with the ability to handle worst case scenario or pay the penalty in time to complete, or lost opportunity.	Must overprovision enough instances to handle worst case scenario or increase operational overhead of scaling up and down nodes as needed. Annual contracts cannot be scaled down.	None.

Elimination of On-premises Related Operational Expense

Eliminating the need to operate hardware on-premises brought about substantial operational savings for the organizations that switched to BigQuery. Deploying the legacy DW solution on AWS infrastructure eliminates the need to manage and maintain hardware, but a node-based AWS solution still introduces some operational overhead versus the truly serverless design of BigQuery.

- **Eliminate infrastructure-related costs –** By deploying your EDW solution in the cloud (AWS or BigQuery), organizations eliminated the need to pay monthly bills to power, cool, and house the EDW hardware.

- **Reduce support and maintenance contracts –** CSP enterprise support contracts are typically less expensive than those required to support the on-premises hardware-based solution, as there is no hardware to maintain. As cloud spend goes up, the Google Enterprise support rate (as a percentage of total cloud spend) is reduced to as low as 3%. Customers deploying the legacy EDW on AWS must manage two points of support and will need to purchase support from both the EDW vendor and AWS.

- **Eliminate time spent performing updates and upgrades –** BigQuery customers reported not having to perform updates and patching for any hardware, software, or security tools. All updates happen non-disruptively behind the scenes at Google. The on-premises solution required disruptive upgrades with a requirement to reboot the nodes several times per year. Updating the EDW software running on the AWS instance, or upgrading to a more powerful instance type also requires manual steps, incurring disruption to the business.

- **Eliminate license management –** The legacy EDW on-premises solution requires management of licenses, unlike cloud-based solutions.

- **Eliminate up-front capacity planning –** Customers revealed this (along with unpredictable performance) as a top reason for switching to BigQuery. They reported painful months of planning, purchasing, installing, testing, and troubleshooting when increasing the capacity or capabilities of their on-premises EDW. While running the legacy EDW solution on AWS instances does reduce some burden of installing and testing hardware, customers must still deploy additional instances, consuming time and resources. BigQuery completely eliminates the need to plan for capacity beyond the number of slots you wish to reserve. The number of slots can be adjusted in hours instead of weeks or months.

> *"We were spending quite a bit of time on capacity planning – this was the biggest reason we needed a change"*

> *"…we were increasingly getting requests to keep data around for an extended period of time, which is costly on [Legacy EDW Vendor] config – data needed to be moved between systems"*

	Legacy EDW On-premises	Legacy EDW on AWS	Google BigQuery
Ongoing Operational Expenses	• Power /cooling /floorspace. • Maintenance contracts. • License management.	None.	None.
Support Contracts	Typical legacy support contracts around 6% of HW/SW cost annually, increases substantially after three years.	• EDW support bundled into subscription. • AWS: Minimum of $15k, as low as 3% of cloud spend.	Minimum of $15k, as low as 3% of cloud spend.
Capacity Planning	• Must plan to grow the capabilities and capacity of the system. • Process takes months of planning, purchasing, installing, testing, and tuning.	• Similar planning process to scale up nodes or move workloads to more powerful nodes. • 1-2 weeks of planning, installing, testing, and tuning,	None. Adjust slots as needed.

Reduction in the Cost of Daily EDW Administration

Daily EDW administration often consists of tasks such as managing the ETL process between systems, configuration, management, and monitoring of the platform and database, troubleshooting, access control, security, develoment of new services, collaboration with business analysts, providing data for reports and dashboards, and application integration.. ESG found that BigQuery can eliminate some tasks altogether, and make other tasks simpler and quicker through automation and integration.

- **Eliminate hardware-centric administrative tasks –** BiqQuery customers saved time by not having to administer the hardware, provision resources up-front, allocate storage, run vacuum processes, etc.

  "We had 2 FTE for "Infrastructure" and 2 FTE for SW Dev (Queries/Scripts/Tools/Transformation, etc.). [After moving to BigQuery] we were able to reallocate the Infrastructure FTEs to other functions that required resources"

- **Eliminate the need to monitor and remediate capacity and performance Issues –** ESG found that administrators do not need to spend time monitoring the capacity and performance of the system and tuning or rebalancing for optimal performance or capacity utilization. For those that do wish to monitor the BiqQuery environment, there are many functions that allow for insight into query performance, trends, billing, and system audits with the ability to create automated alerts based on thresholds.

  "Demand was growing and was not going to stop - We were having a hard time managing that in our on-premises environment"

- **Simplify ETL process –** The customers that ESG spoke with reported that there were flexible options available to them that allowed 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.). These flexible options allow for greater possibilities and capabilities through the use of other Google products.

- **Simplify legacy EDW related tasks –** BigQuery customers reported saving time by eliminating backups and system maintenance tasks, simplifying onboarding, managing workload priorities, maintaining partitions, etc.

- **Achieve greater flexibility of SQL queries and schemas –** BigQuery often does not require any schema change from most existing DW solutions, but also provides for some enhanced functions that can help optimize schemas and make queries faster and more powerful.

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

  "Our teams on-premises that run our datacenters, they do a great job…but that's a large investment and a lot of time"…"we didn't want to have all the stress of the DevOps"

	Legacy EDW On-premises	Legacy EDW on AWS	Google BigQuery
Legacy Hardware-centric EDW Administration	Manage nodes, allocate and provision resources, run cleanup processes, backup/restore, etc.	Manage nodes, allocate and provision resources.	None.
Monitoring / Troubleshooting / Tuning	Correlation between multiple interfaces and technologies, tune capacity and performance of nodes.	No hardware issues or interfaces, but node-level tuning may be required.	None required, but advanced monitoring and reporting is available.
Legacy EDW Tasks	• On-premises ETL. • Tasks to maintain and protect system. • Rigid schema, inflexible options. • Maintain user access and security.	• On-premises or cloud ETL. • Tasks to maintain and protect system. • Rigid schema, inflexible options. • Maintain user access and security.	• On-premises or Cloud ETL. • Flexible, optimized schema. • Maintain user access and security.

Automated Scalability, Availability, and Agility

A direct comparison between the legacy EDW on-premises solution and BigQuery is difficult as the two solutions often do not provide equivalent functionality. ESG learned that after organizations made the switch to BigQuery, they found that they could achieve things that they never would have been able to do using the legacy on-premises solution. This enriched functionality leads to other benefits that directly impact business operations and further increase the value of BigQuery as a next-generation EDW solution.

- **Ability to scale up or down to handle peaks and valleys –** Customers reported that once other areas of the business began to expand beyond traditional queries, they could easily scale up the number of slots to increase the organization's capabilities. Providing resources to handle new projects would take months of planning for a legacy on-premises solution, resulting in missed business opportunities and revenue streams.

    "Two or three years prior - capacity expansion took five months of planning and required a three-month outage. And we used up all that capacity in about a year to two years. It was time to do another one. "

- **Improved query performance/faster insight –** Query performance was a top concern for organizations relying on their legacy on-premises EDW solutions. Poor query performance was the number one reason for one organization's move to BigQuery. BigQuery leverages massively parallel processing while continuously optimizing algorithms and data layouts, resulting in faster and more predictable query time.

    " You will always have people asking for faster execution no matter how fast you go, but the number of instances of customers complaining about time to complete queries or reports is very minimal now. Before it was maybe five instances per week, now it's fewer than one for sure."

- **Ability to quickly act on new opportunities** – Faster queries and 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. BigQuery customers reported that they were much more agile and able to act

on opportunities much faster with no impact to the business. Organizations relying on legacy EDW solutions will miss out on these opportunities or arrive late to the party once resources have been allocated to satisfy new workloads.

> *"We had a ton of workloads that we couldn't even run in our on-premises environment because we just didn't have the capacity on there"*

- **Cross-functional organizational efficiency** – Organizations reported that they were impressed with how BigQuery helped to make operations run more smoothly. Organizations noted improved ETL processes, the benefit of self-service, and the abilities to manage access using the organizational hierarchy, quickly enable new functions within the organization, simplify billing and reporting, limit and share resources across divisions, and seamlessly integrate their EDW with other data services. While each of these benefits can be accounted for in areas already discussed, BigQuery improves and streamlines the entire EDW process, resulting in a more efficient organization.

> *"We decided if we are going to do this… we were going to modernize everything…In short - it is game changing, and all the enhancements that are coming in - are making it even better for us"*

Migration Considerations

Migrating on-premises EDW workloads to BigQuery 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 BigQuery. While ESG's model built these costs into our analysis in good faith based on blended reports, actual costs will vary dramatically depending on your current environment, plan of execution, and additional capabilities you wish to add.

- **Migration strategy for EDW** – Organizations should consider if 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 re-tooling 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, include 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/or SaaS licenses, but should not forget to balance these new costs against any potential savings gained by migrating to BigQuery.

> *"I considered that this project was done in a record short period of time. I am proud of my team - my team did an amazing job…We are all very proud of what we accomplished"*

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 legacy on-premises and cloud-based EDW solutions. ESG's interviews with customers who have recently made the transition from on-premises to BigQuery, combined with 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 in which a large organization (over 10,000 employees) had deployed an on-premises EDW using a previous generation solution supplied by the legacy EDW vendor. The organization historically had supported hundreds of databases and sources of data collected from multiple locations across the US, but the solution no longer met their performance and capacity demands. The organization operated with about 380 TB of usable data in their EDW and was looking to migrate this data to a solution that would offer greater performance and flexibility. The organization was rapidly growing its analytics capabilities and planned to rapidly expand both the compute and storage capacity annually. The initial deployment would consist of 24 compute nodes and 380 TB of usable storage. Another 12 compute nodes and 190 TB of usable storage would be added in year 2, followed by 8 more data nodes and 128 TB of storage in year 3.

ESG estimated the costs for migrating the existing data and satisfying the EDW requirements described above to each of three options: 1) the latest all-flash-based version of the on-premises EDW solution, 2) the cloud-based EDW software running on AWS IaaS (provided by the same legacy EDW vendor), and 3) Google BigQuery. ESG created and leveraged detailed costing models for each of the categories described in the table below to satisfy the requirements of the modeled organization.

	Legacy EDW On-premises	Legacy EDW on AWS	Google BigQuery
Up-front Capital Investment	• Cost of hardware. • Cost of software. • Maintenance /support contracts.	• AWS compute (one-year all up-front RI). • One-year annual EDW license and support.	• None.
Cost of Deployment and Migration	• Capacity planning. • Node planning. • Ordering, purchasing, etc. • On-prem data movement. • Schema updates. • Feature enhancements. • Dev/test.	• Capacity planning. • Node planning. • Move data to the cloud. • Schema updates. • Feature enhancements. • Dev/test.	• Plan # of slots (no risk). • Move data to the cloud. • Schema updates. • Query updates. • Retooling. • Dev/test.
Monthly Cloud Costs	• None.	• AWS storage costs. • Supporting cloud services. • AWS support	• BigQuery flat-rate pricing. • Cost of storage. • Supporting cloud services. • Google support.
Administrative Costs	• Hardware administration. • EDW/DB administration. • Query administration. • User/process admin.	• AWS administration. • EDW/DB administration. • Query administration. • User/process admin.	• Query administration. • User/process admin.
Operational Costs	• Power. • Cooling. • Floorspace.	• None.	• None.

Figure 2 shows the estimated three-year total cost of ownership for each of the three solutions. When estimating these costs, we accounted for discounts that both the legacy vendor (hardware, software, support, and annual agreement discounts) and Google (flat-rate pricing) can offer its enterprise customers.

Figure 2. Estimated Three-year Data Warehouse Solution Total Cost of Ownership (TCO)
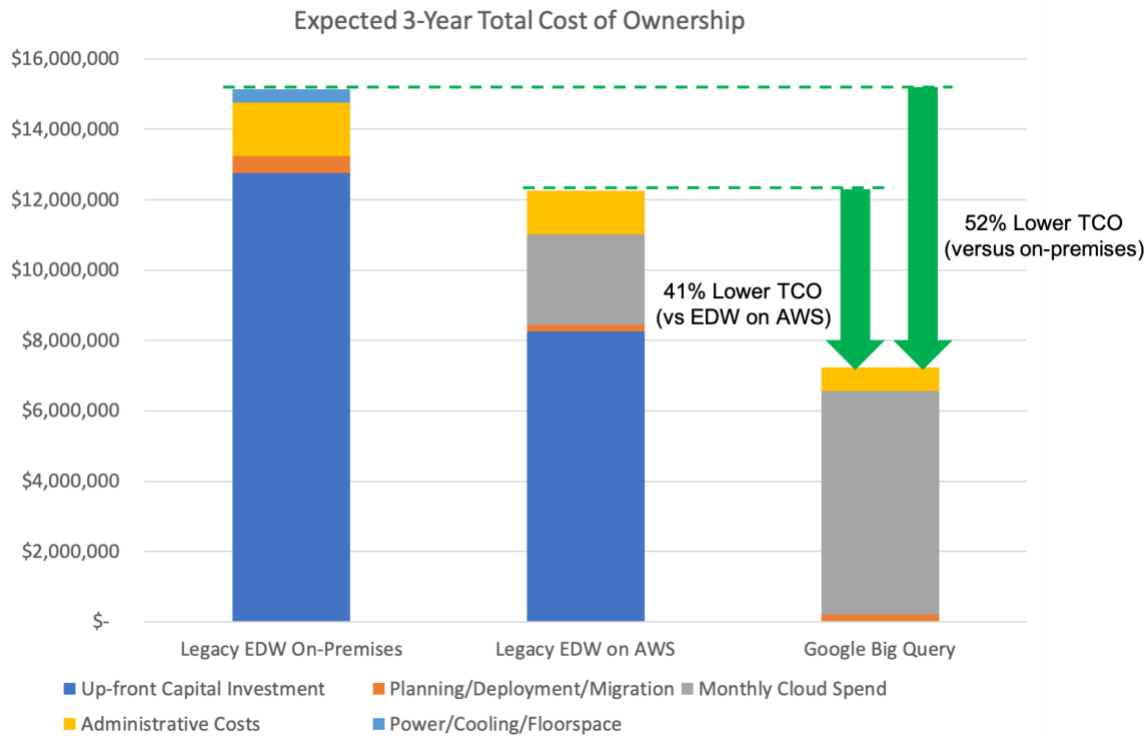
Not surprisingly, the on-premises EDW solution had the highest expected TCO of the three deployment options ($15.1M), followed by the EDW software running on AWS infrastructure ($12.3M). Google BigQuery ($7.2M) provided the lowest expected three-year TCO at a savings of 52% versus the on-premises EDW and 41% versus the legacy EDW on AWS infrastructure.

Up-front Capital Investment

To migrate the existing EDW to the latest release of the on-premises EDW solution, we assumed that the organization replaced its existing hardware nodes (comprised of compute and storage capacity) with an equivalent number of active nodes (nodes that house the databases) from the latest solution release. Over our analysis period, the number of nodes and storage capacity grew to support a greater number of business functions. The organization also had to purchase passive nodes to protect against hardware failure of the active nodes (ESG assumed one passive node for every 12 data nodes in a cabinet).

ESG found that the legacy EDW solution required a significant up-front capital investment while Google BigQuery required zero up-front investment.[3] Much of the up-front investment would be paid in year 1, but some would be paid at the start of each subsequent year (growth and annual agreements). ESG priced the required hardware ($4.7M), software ($4.4M) and support ($3.6M) for the on-premises solution conservatively. ESG also noted that software costs can accumulate quickly, depending on organizational requirements. When purchasing the required software to operate the on-premises

---

[3] ESG's analysis did not take into consideration any on-premises hardware purchases that would be required for staging the ETL process, assuming these costs would be equivalent across the three solutions. If required by your BigQuery implementation, there may be some up-front investment.

EDW, the organization must purchase both the operating system and the base software for each active and standby node. For database features that are not available in the base software, the organization must purchase individual licenses for each node as well (e.g., data security, system management, etc.).

To provide the largest savings available to customers for EDW on AWS, ESG chose to price the options for paying all up-front annual agreements for both the EDW software and maintenance (23% discount versus hourly) and AWS EC2 reserved instances (RI) (42% savings over hourly). ESG did not choose further EC2 discounts provided by a three-year RI price, as hardware must be added in each year, and to keep the annual agreements in lock-step. The expected up-front investments are shown in Figure 3.

Figure 3. Estimated Up-front Investment (Paid at Time of Purchase or Via Annual Contract)
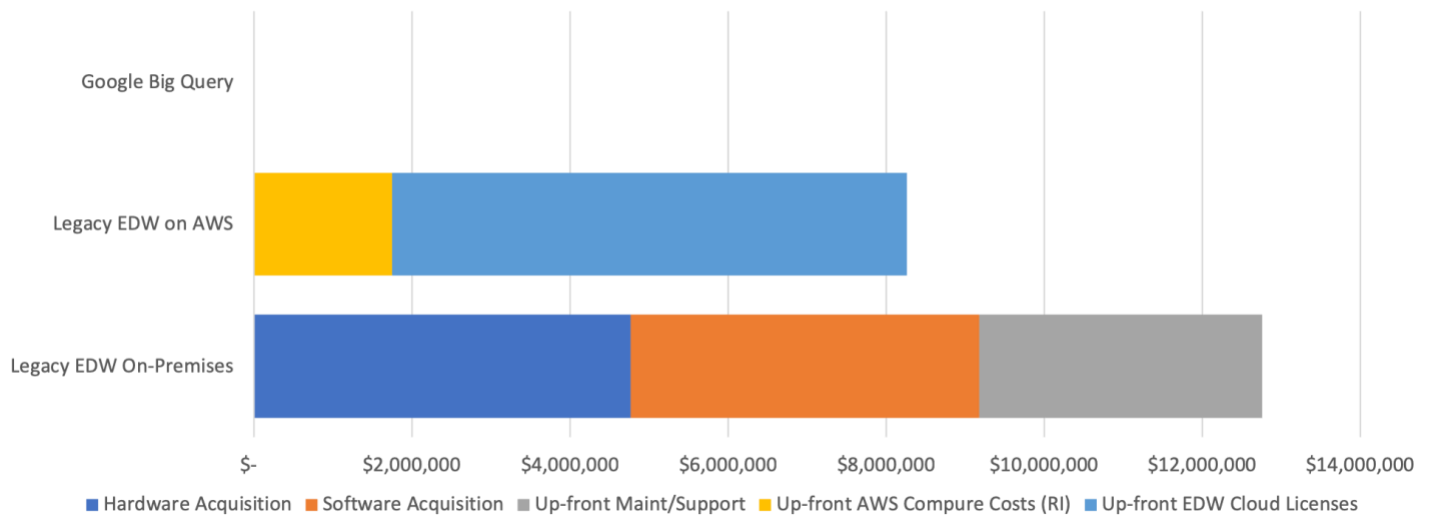
Cost of Planning, Deployment, and Migration

ESG considered the expected time spent to plan for, purchase, and deploy the required hardware, software, support, and services for each of the deployment options along with the number of internal and/or professional resources that would be required to take part in the activities. ESG assumed a cost of $50 for each internal man-hour and $250 for each professional service man hour required. Costs were an analyzed for both the initial deployment and for annual expansion. The EDW deployments (both physical and in the cloud) involved time to correctly size and deploy the nodes and storage and rebalance and tune where needed. After initial migration, BigQuery planning and deployment only involved growing the number of slots when needed. Figure 4 compares the expected costs for each solution for planning, purchasing, deployment, and installation.

Figure 4. Estimated Cost of Planning, Purchasing, and Hardware Deployment
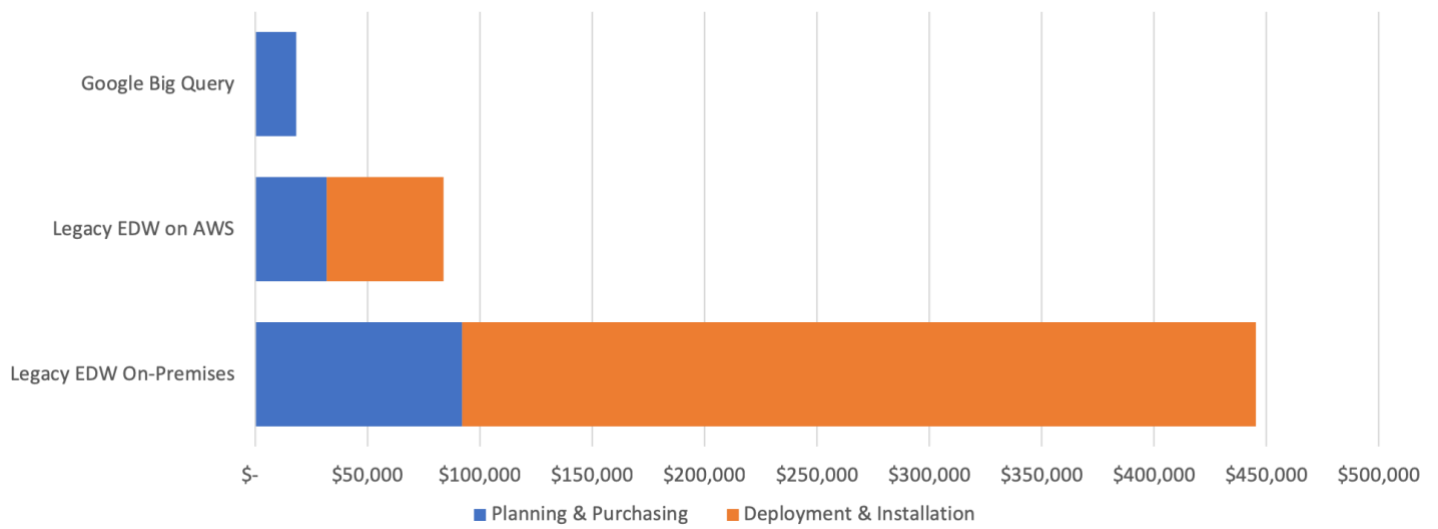
As expected, the one-time cost of migration to BigQuery required a larger investment than did migrating to the legacy EDW on-premises or on AWS. ESG modeled the expected costs of the actual physical transfer of data, schema updates, feature enhancements, development and test, and training both administrators and end-users in the new technology and processes. ESG based the assumptions of man hours required for these stages on a blended result of times reported in customer interviews. We assumed the modeled organization has a team of two general IT admins to handle the data movement and four developers to handle development and test work (schema updates, query modifications, feature enhancements, dev/test). Figure 5 shows ESG's estimated cost of migration.

Figure 5. Estimated Cost of Migration from Gen-1 Legacy On-premises EDW

Note: One-time investment for change in technology is more than offset by savings and benefits.
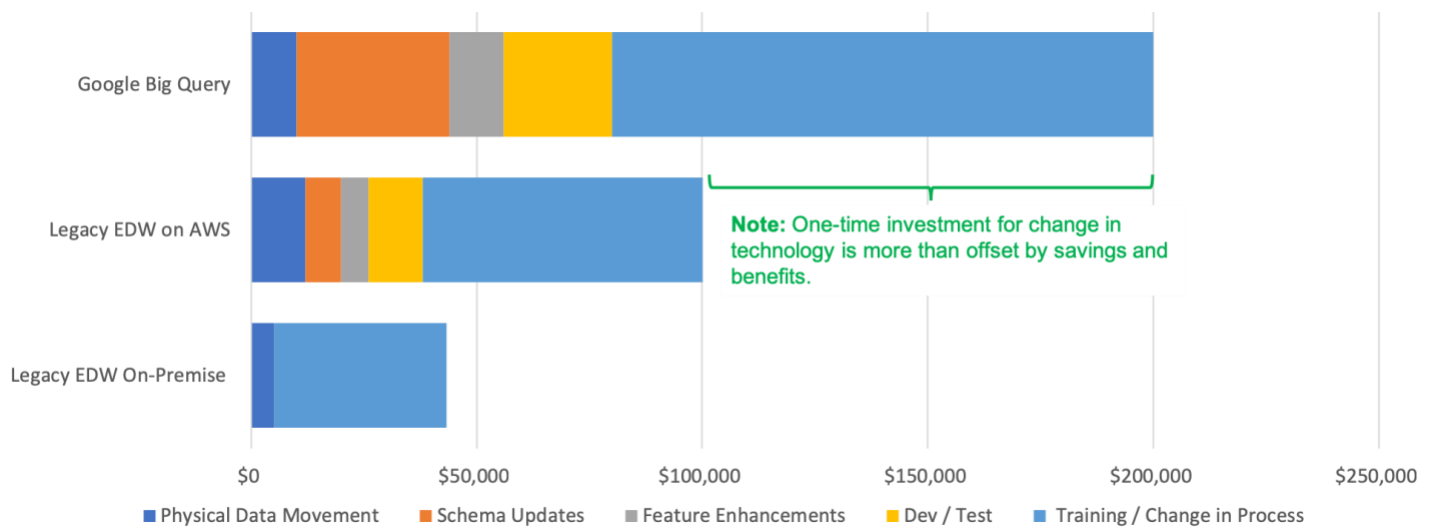
The bulk of the one-time costs depicted in Figure 5 results from migrating code from the legacy EDW to BigQuery and from training staff in the new environment. This additional investment is to be expected when doing a migration to cloud and making a change in a vendor, but the reader should notice that this cost is more than offset by a total cost of manpower (deployment, migration, and administration) that is 1.6X to 2.2X lower than the legacy EDW solution options.

Monthly Cloud Costs

Cloud costs for AWS consist of costs paid for software, licenses, and support to the EDW vendor, and compute and storage costs paid to AWS. ESG sized the on-premises EDW on AWS using m4.16xlarge EC2 nodes (to match up best with the vendor-provided performance rating of the deployed physical nodes) and EBS storage (at $0.12/GB/month) as described by the requirements on the AWS marketplace. To provide the most advantageous costs to the AWS EC2, we assumed that all costs were paid up front annually rather than monthly. In Figure 6 below, these up-front costs are amortized as monthly costs depicted as patterned boxes for illustrative purposes.

BigQuery cloud costs (three-year cumulative spend of $6.3M) consist of increasing annual slot counts of 5,000 (Y1), 7,500 (Y2), and 10,000 (Y3) as well as BigQuery cloud storage costs. Active data stored on Google cloud storage is billed at $0.02/GB/month and any cold data (data that was not used in the last 90 days) is automatically billed at a 50% discount as long-term storage charged at $0.01/GB/month. ESG assumed that 50% of all data would be active, and 50% would qualify for long-term storage. The breakdown of the expected (normalized) effective monthly cloud spend is shown in Figure 6.

Figure 6. Estimated Monthly Cloud Spend (Up-front Costs Shown are not Paid Monthly)
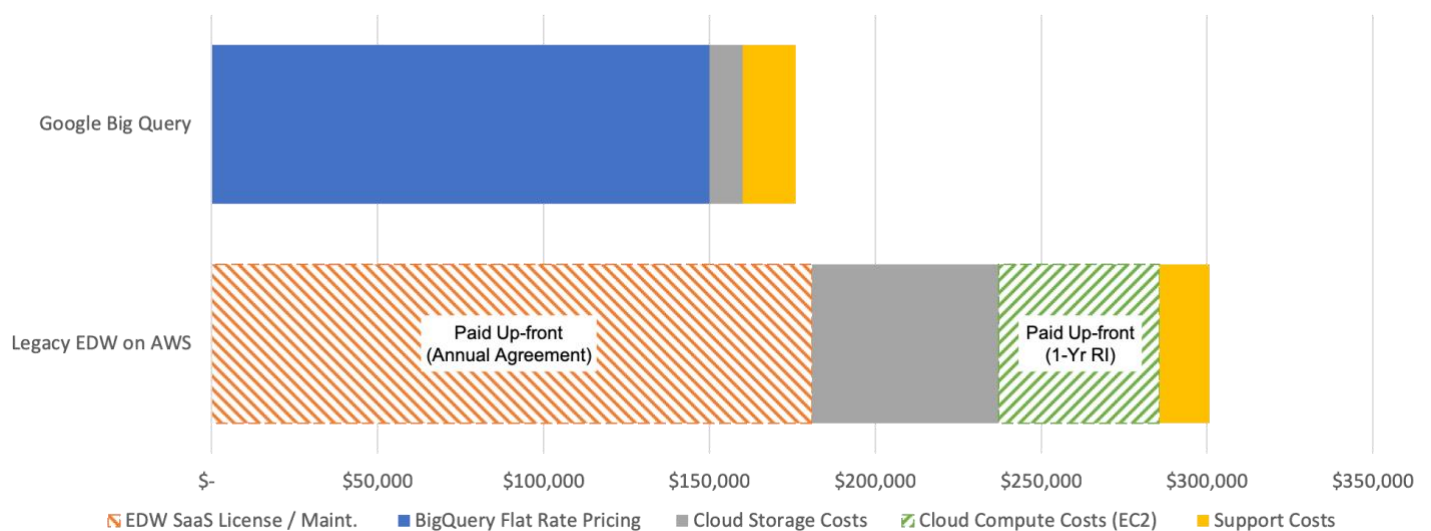
Source: Enterprise Strategy Group

## Administrative and Operational Costs

The costs required to operate the EDW hardware, including power, cooling, and data center floorspace, were calculated based on the published specs and an average cost of electricity of 13 cents/ kWh and a floorspace cost of $12K per 42U rack per month resulting in a three-year expected operational cost of $387K.

ESG modeled the expected daily administrative costs of managing each of the three solutions over a three-year period in each of five administrative categories. The on-premises EDW solution required hardware administration, which included time to monitor, maintain, update, replace, and configure servers, storage, and networking resources. The EDW solution on AWS required roughly half of this effort to monitor, maintain, configure, tune, and perform billing and licensing tasks on virtualized compute instances. Both of the EDW solutions required a similar effort to administer the EDW software and databases. In contrast, BigQuery is completely serverless and required none of this administrative overhead.

For ETL and Query administration, ESG modeled the time required to manage the batch ETL process and manage queries and generate reports. While some of the cloud tools helped lower this cost for the EDW on AWS over the on-premises solution, BigQuery provided further savings based on the fact that queries can be written simpler and can be optimized to

execute faster, and there is more opportunity for automation and self-service capabilities. The breakdown of the estimated three-year costs to administer the three DW solutions are shown in Figure 7.

Figure 7. Estimated Three-year Administrative Costs
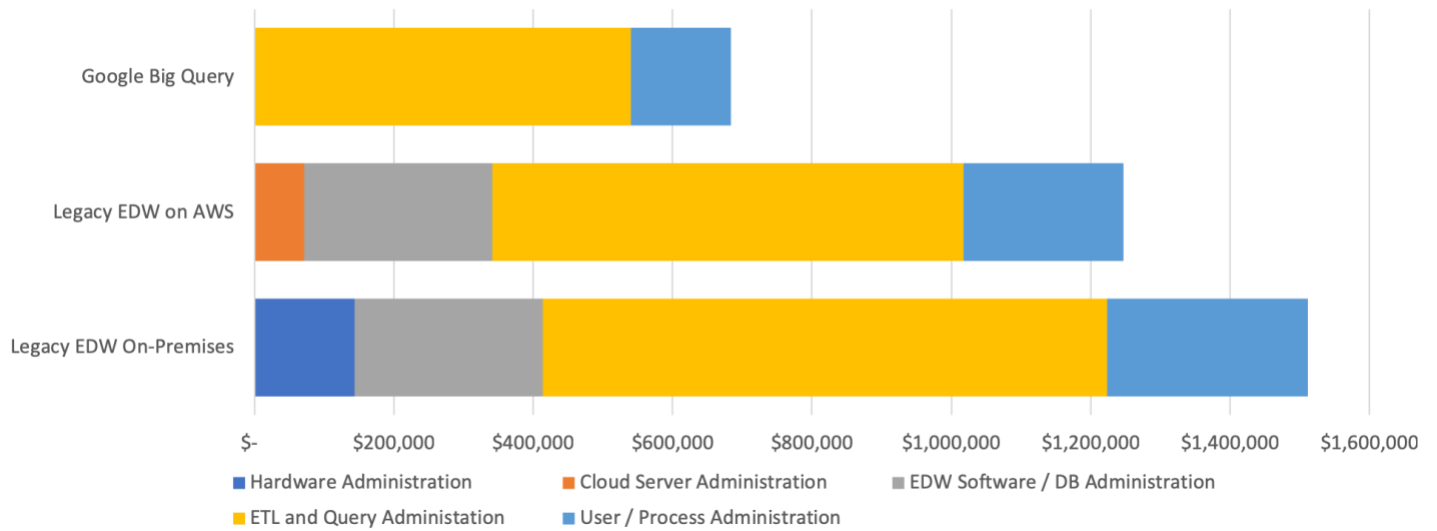
Source: Enterprise Strategy Group

Additional Savings and Benefits

In addition to the TCO savings summarized above, ESG notes that a typical large organization such as the one modeled here could potentially realize millions of dollars in additional savings and benefits brought about by the operational agility of BigQuery. BigQuery provides the ability to scale up and down to address new or changing business opportunities, delivers improved query performance for faster insight and competitive advantage, and enables cross-functional operational improvements and seamless integration with other cloud services. When all of this is taken into consideration, the opportunity for additional savings and benefits are astounding. Through the newfound ability to provide more comprehensive intelligence to the organization in a faster and more agile manner, large enterprise organizations could easily expect overall savings and benefits in the millions to tens of millions of dollars by migrating from on-premises EDW solutions to Google BigQuery.

The Bigger Truth

Legacy on-premises EDW solutions have served as the backbone for business intelligence for decades. The insight produced justified the extraordinary up-front capital investment and ongoing operational costs. As new streams of information are made available and consolidated from a variety of internal and external sources, business requirements change dynamically, yet the operational rigidity of traditional EDW solutions are proving to be a roadblock to modernization, ultimately placing organizations at a competitive disadvantage. 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.

Once a leader in EDW deployments, the legacy EDW vendor referenced here is increasingly hampered by the constraints of offering an on-premises-centric EDW solution. While running virtualized instances of the legacy solution on IaaS certainly removes the burden of managing and maintaining physical hardware, the limitations of managing and maintaining a legacy node-based EDW remain. Google BigQuery is designed to remove all of the physical and logical burden of managing, monitoring, and maintaining EDW infrastructure, allowing organizational resources to focus on obtaining intelligence rather than the process of generating it.

ESG validated the cost savings and operational advantages that organizations have seen since migrating their legacy EDW implementations to Google BigQuery through a series of interviews with the organizations and their supporting customer engineering teams. The information collected through these intensive interviews helped form the basis for the model's assumptions around such factors as the size and costs of deployments, operating costs, administrative man hours required, time and effort spent to migrate to BigQuery, and the operational benefits seen since making the switch to BigQuery, among others. Using the knowledge gained in these interviews, ESG created a three-year modeled analysis that shows that organizations can save up to 52% by using BigQuery over an on-premises EDW and up to 41% over an EDW deployed in the cloud.

Your organization may have made large investments in legacy, on-premises EDW solutions in the past, but this technology may no longer provide the business advantage and operational flexibility that you need from your enterprise data warehouse. The infrastructure, people, and processes are most likely well established, and the thought of disruption brought about by change may be intimidating. However, those that fail to adapt to today's insight-driven world risk being placed at a significant competitive disadvantage. One of the customers involved in the study had migrated one of the largest legacy on-premises EDW deployments in the United States to BigQuery. Not only did BigQuery meet the needs of the organization, but it was also able to easily scale up and extend the capabilities of the organization beyond what could have been achieved with an on-premises EDW.

> *"We did significant proof of concepts on other cloud-based analytics options - Google came out in front as the most agile, flexible, and scalable solution"*

If your organization is struggling to overcome the burden of a legacy on-premises EDW, it is certainly time to invest in a more agile solution. ESG's analysis shows that Google BigQuery offers a lower expected TCO, substantially streamlines operations, and delivers far more agility and flexibility for bringing the value of next-generation actionable insight across the 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.