Enterprise Strategy Group | Getting to the bigger truth.™

Technical Validation

Achieving Operational Simplicity and Optimal Price-performance with Actian Avalanche Cloud Data Platform

By Alex Arcilla, Senior Validation Analyst; and Tony Palmer, Principal Validation Analyst

April 2022

Introduction

This ESG Technical Validation highlights how Actian Avalanche cloud data platform enables organizations to execute faster iterations on data models to support real-time decision making for greater business intelligence. A combination of hands-on analysis and customer use cases were used to validate the advantages of leveraging Actian Avalanche cloud data platform for robust advanced analytics.

Background

Collecting and using data in real time is transforming and empowering organizations. A growing variety of sources generate real-time data from devices and machines, customers, suppliers, partners, and market interactions. Messaging applications are now real-time, and sensor-enabled machines deliver constant streams of data. Social media delivers real-time feedback and insights to consumers, and clickstream data from digital commerce can deliver predictive value to companies. All of these data sources present the opportunity to add significant business value. Subsequently, organizations want to use this data to understand customer needs, identify key trends, and design better products, as well as to prevent, mitigate, and solve problems. As a result, real-time data analytics has become a key business priority. In fact, when asked which business initiatives they believed would drive the most technology spending in their organizations in 2022, 33% of respondents to ESG's Technology Spending Intentions Survey identified improving data analytics for real-time business intelligence and customer insight as a top business initiative, making it the second most-cited initiative, behind strengthening cybersecurity (see Figure 1).[1]

Figure 1. Top Five Business Initiatives Driving Technology Spending

Source: ESG, a division of TechTarget, Inc.

Actian Avalanche Cloud Data Platform

Actian Avalanche is a modern cloud data platform designed to deliver secure access to all the disparate types of data that users require to effectively perform their jobs. By employing business intelligence, visualization, reporting, and machine learning tools, users can easily access diverse data sources to gain deep insight and business advantage—without IT intervention.

[1] Source: ESG Research Report: *2022 Technology Spending Intentions Survey*, November 2021.

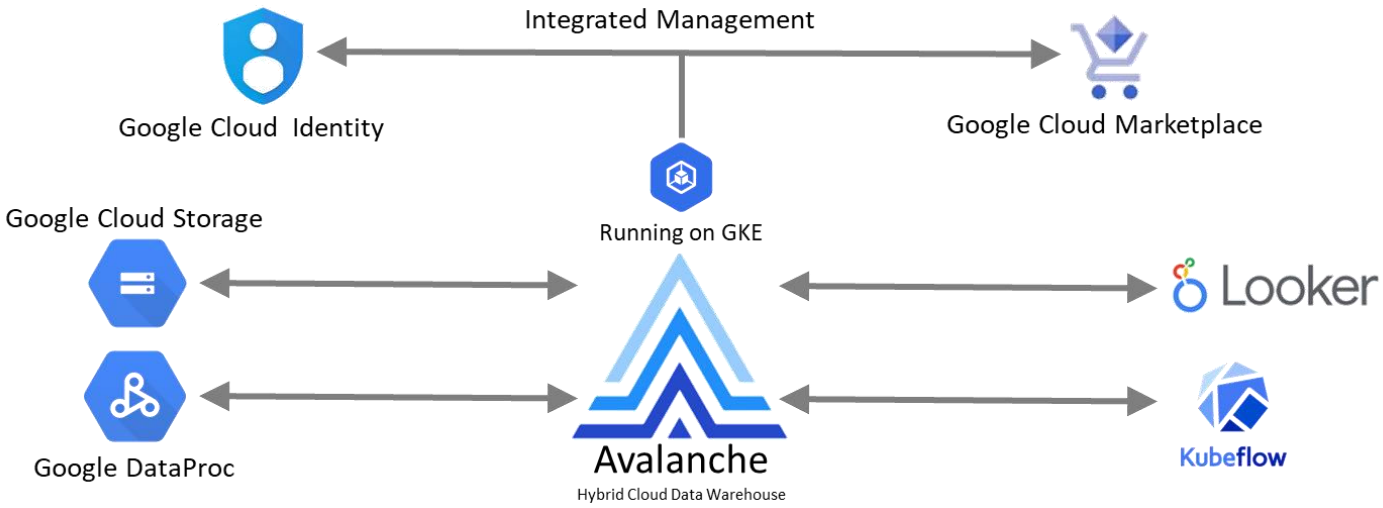
Actian and Google Cloud have built a strategic partnership to deliver business intelligence and operational analytics both in the Google Cloud and on-premises via Google Anthos with price-performance top of mind.

The Avalanche platform is offered through Google Marketplace as a fully-managed service billed on the basis of resources used. The Avalanche platform is designed for performance, scalability, and flexibility to meet the needs of organizations of any size. In addition to availability on Google Cloud, the Avalanche Cloud Data Platform also supports Amazon Web Services and Microsoft Azure environments. For the purposes of this report, we will focus on the Avalanche platform on Google Cloud.

The analytical engine in the Avalanche platform is built to deliver sub-second responses to user queries and associated analytics that span billions of rows. It uses vectorized processing, CPU cache capabilities, and creative use of columnar storage to deliver high performance processing power.

The implementation of the Avalanche platform on Google Cloud enables it to take full advantage of the platform's built-in elasticity so the Avalanche platform can automatically increase or decrease the number of processing cores and the amount of RAM required to meet each user's performance requirements. The Avalanche platform is usage-based; it increases or decreases the amount of storage allocated to datasets and applications dynamically, eliminating the need to increase the number of servers required to manage storage.

Figure 2. Actian Avalanche



Source: ESG, a division of TechTarget, Inc.

The Avalanche platform enables users to easily interact with structured, unstructured, and semi-structured datasets—from on-premises to cloud-based solutions, persistent as well as in motion—while using their preferred tools and techniques. In addition, the platform provides direct in-system support for Looker and Google Analytics ecosystem tools as well as their equivalent non-Google counterparts.

ESG Technical Validation

ESG performed evaluation and testing of the Actian Avalanche Cloud Data Platform via demos conducted at Actian's facilities in New York, NY. Testing was designed to demonstrate the ease of configuring the warehouse and data integrations, the flexibility of using common tools already used by enterprises for data-related activities, and the optimal price/performance that end-users can gain.

Platform Ease of Use and Flexibility

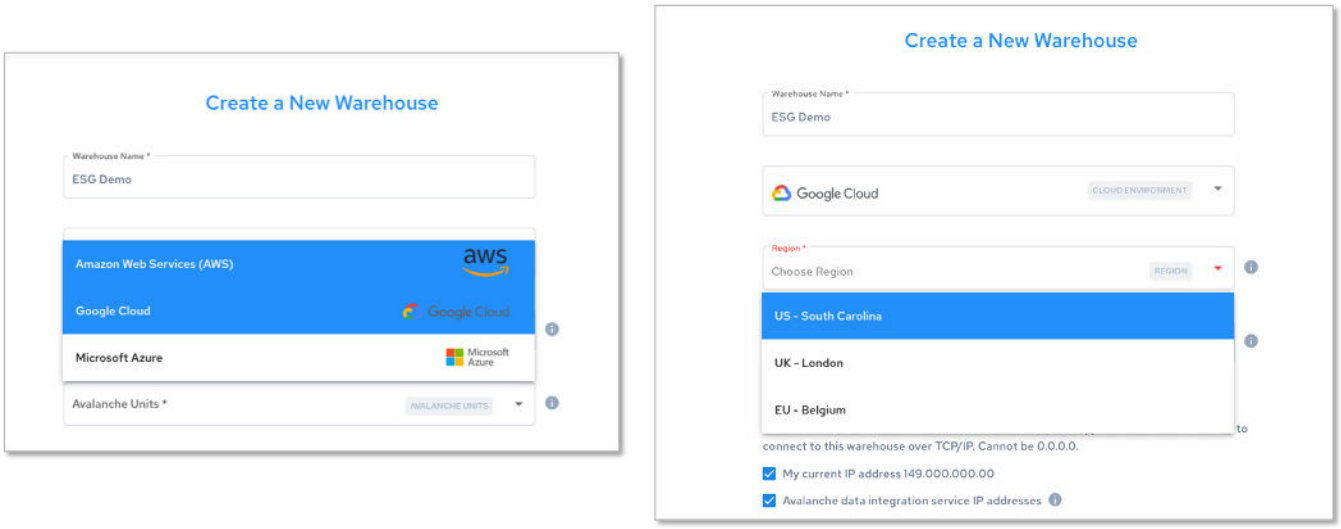
When deploying cloud-based data warehouses, architects can opt to size the appropriate number of compute instances to support a data warehouse, but the time and effort spent may interfere with actually using the data in a timely manner. In addition, as business needs change, scaling or sizing down the warehouse typically is a manual process since end-users must closely monitor usage and add or remove instances as appropriate. Should compute resources go unused, money is wasted. On the other hand, organizations can leverage the Avalanche Cloud Data Platform to simplify the deployment of a data warehouse while minimizing expenses.

ESG Testing

ESG first navigated to the Avalanche platform GUI and clicked on ***Create Warehouse*** (see Figure 3). While the report focused on the integration of The Avalanche platform and Google Cloud, we noted that organizations had the flexibility of using Amazon Web Services and Microsoft Azure.

After naming the warehouse "*ESG Demo*," we chose Google Cloud and the region in which the warehouse would reside. (Different regions would be displayed based on the chosen public cloud.) We noted how this would be important to consider when it came to issues such as data locality and data sovereignty. To ensure that only authorized users could access the data, IP addresses or CIDR blocks could be added under the ***IP Allow List*** section. We noted how easy it was to specify who could access the warehouse.

Figure 3. Configuring Data Warehouse on Google Cloud

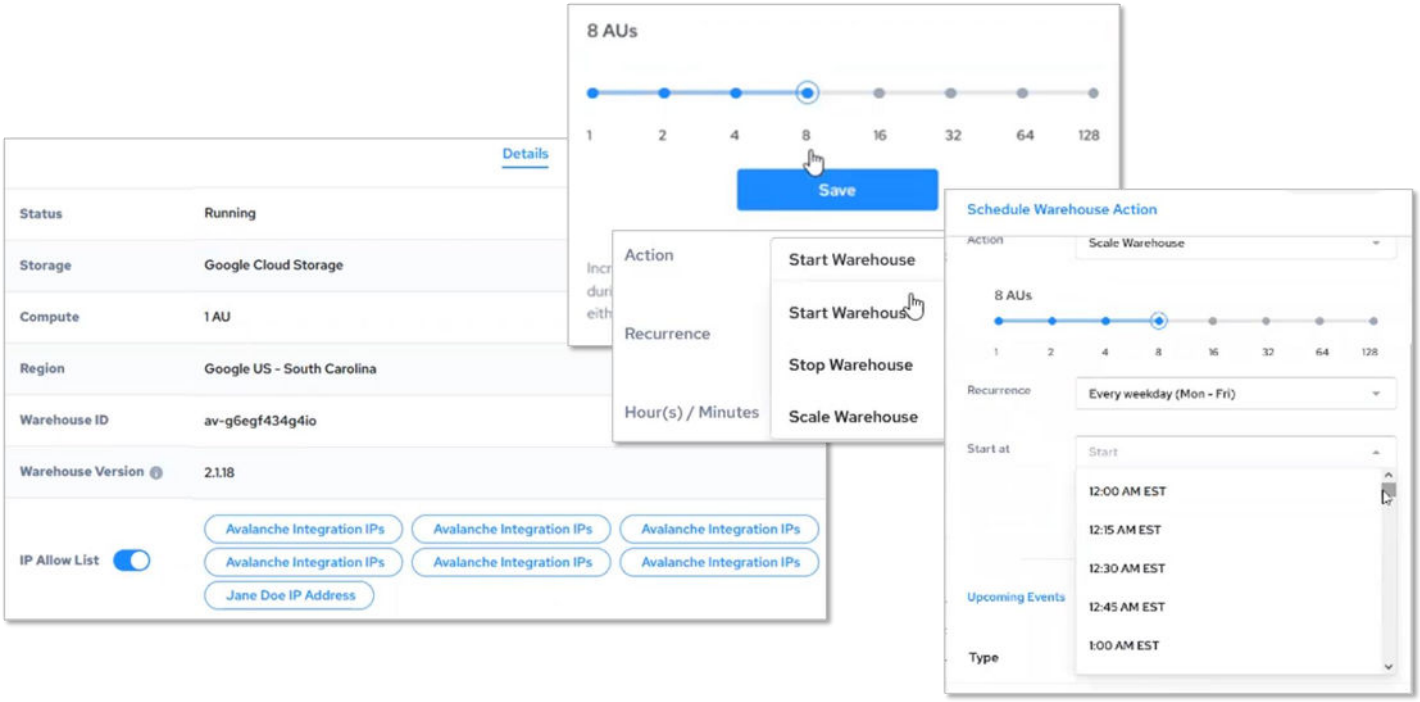


Source: ESG, a division of TechTarget, Inc.

Once we created the warehouse, we were presented with the basic configuration (see Figure 4). At this point, we were presented with options to customize. One critical option was choosing the size of the warehouse based on Avalanche Units (AUs). An AU is a logical measure of computing cluster capacity used by end-users to determine the total amount of

processing power and storage required for the desired data analytics workload.[2] The default size of a newly created warehouse was four AUs; clusters can scale up to 128 AUs.

Figure 4. Customizing the Data Warehouse



Source: ESG, a division of TechTarget, Inc.

To maximize the use of the cluster while minimizing expenses, ESG saw how we could schedule times in which the cluster scales due to less usage (such as off-hours) and for planned events (such as closing out end-of-quarter activity). We could select days that the cluster would resize using the ***Recurrence*** drop-down menu, then specify start and stop times that the cluster would add or remove AUs during those days.

Why This Matters

ESG research has found that 32% of respondents consider increasing employee productivity as one of their top five criteria for justifying IT investments to their business management teams in 2022.[3] Setting up cloud-based data warehouses to support analytics workloads is not a straightforward exercise since organizations need to determine the most appropriate amount of both compute and storage capacity to handle the anticipated number of users, queries, and analyses. Exacerbating the issue is the inconsistency regarding instance types, prices, and services across all public cloud providers, making it more difficult to choose the right amount of compute and storage capacity.

ESG validated that Actian's Avalanche Cloud Data Platform enables organizations to configure data warehouses in a consistent and repeatable manner, thus helping organizations to be more efficient in meeting business needs with minor delay. We verified the ease with which end-users can configure the data warehouse by choosing the required number of Avalanche Units, which removes the need to choose and size the appropriate compute instances. ESG also confirmed the flexibility in scaling the compute cluster during specific time periods on select days, helping organizations to control cloud expenses.

[2] Current pricing is $2.50 USD per Avalanche units per hour.
[3] Source: ESG Research Report: *2022 Technology Spending Intentions Survey*, November 2021.

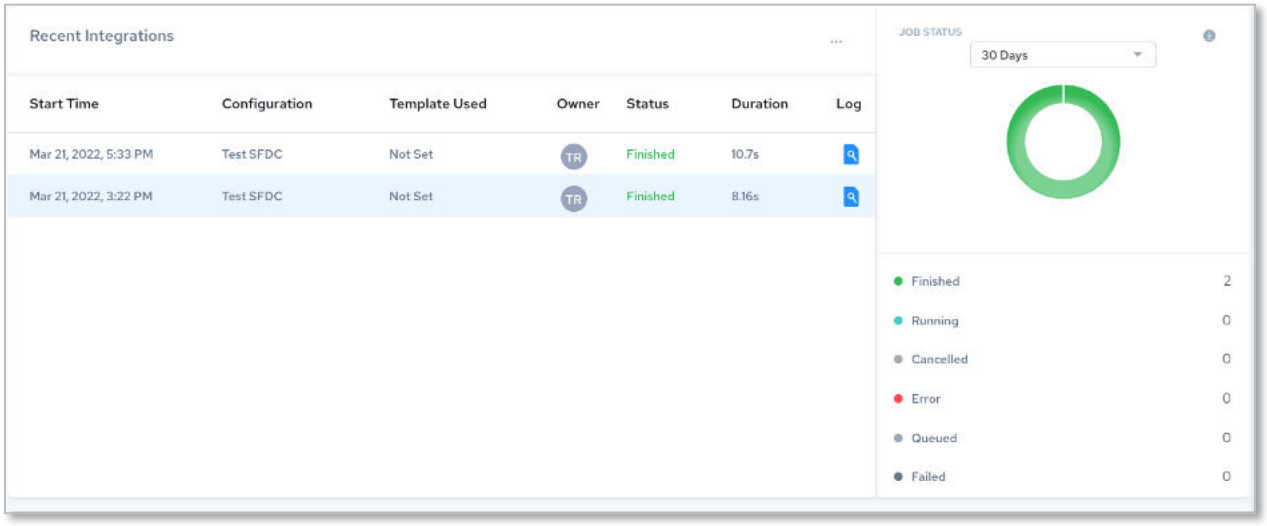
Availability and Flexibility of Data Integrations

Consolidating and preparing disparate datasets from multiple systems, either on-premises or cloud-based, then connecting end-users to those consolidated datasets is an issue faced by organizations every day. For those choosing to employ cloud-based data warehouses, public cloud service providers continue to lack the tools that gather, construct, and normalize datasets. If organizations are unable to provide a complete dataset via the cloud-based data warehouse, the ease of gaining valuable business insights is limited at best.

ESG Testing

ESG first reviewed how to connect third-party application and data repositories to our newly created Actian Avalanche data warehouse. When navigating to the Integrations page, we see dashboard of recently configured integrations, status, and log files.

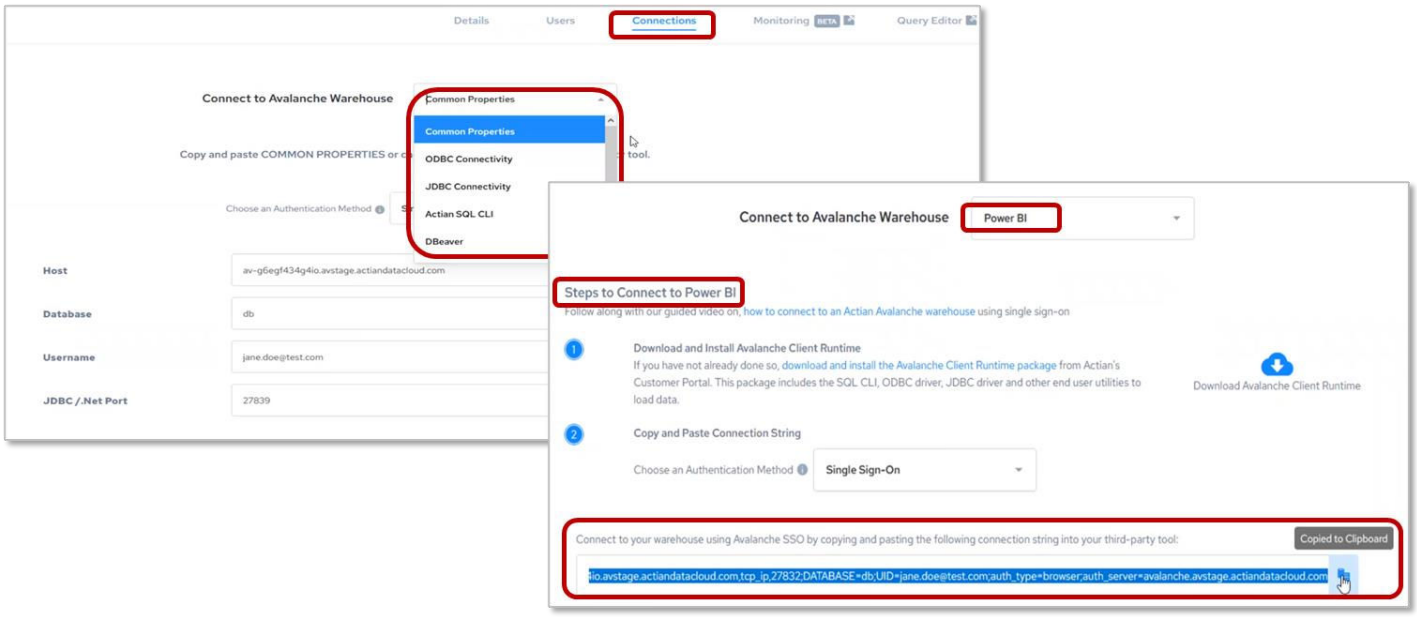
Figure 5. Avalanche Integrations



Source: ESG, a division of TechTarget, Inc.

From the drop-down menu on the ***Integrations*** page, we saw commonly used tools such as Salesforce, IBM Netezza, and Tableau (see Figure 6). After selecting Power BI, we found that the Avalanche platform provided simple instructions for connecting to the data warehouse. In the case of Power BI, we were presented with a connection string to copy and paste in the Power BI tool to upload the desired data.

Figure 6. Choosing a Data Source to Connect with Actian Avalanche Data Warehouse

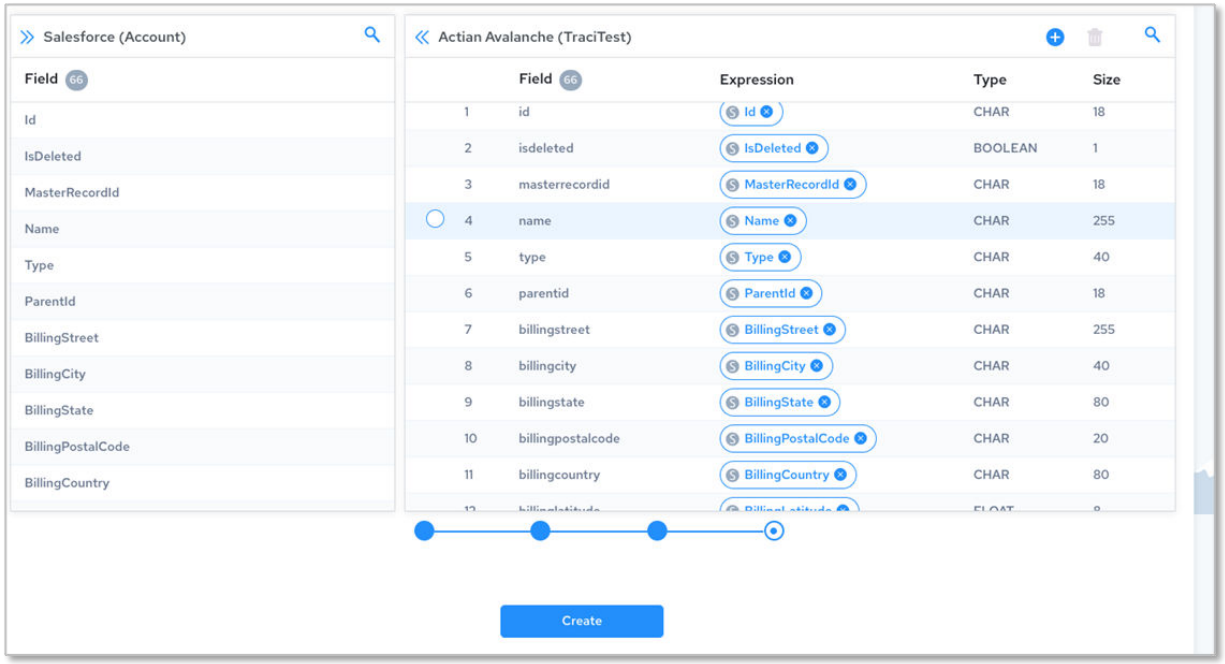


Source: ESG, a division of TechTarget, Inc.

ESG noted how the Avalanche platform enables organizations to use the tools that they already use for data management and analysis. Organizations are not forced to migrate all data into the Actian Avalanche Cloud Data Platform. End-users can also bypass traditional extract-transform-load (ETL) tools.

When connecting to a data source, the Avalanche platform automatically maps the source data to the table columns to save time, as seen in Figure 7. This can be adjusted by the user as needed.

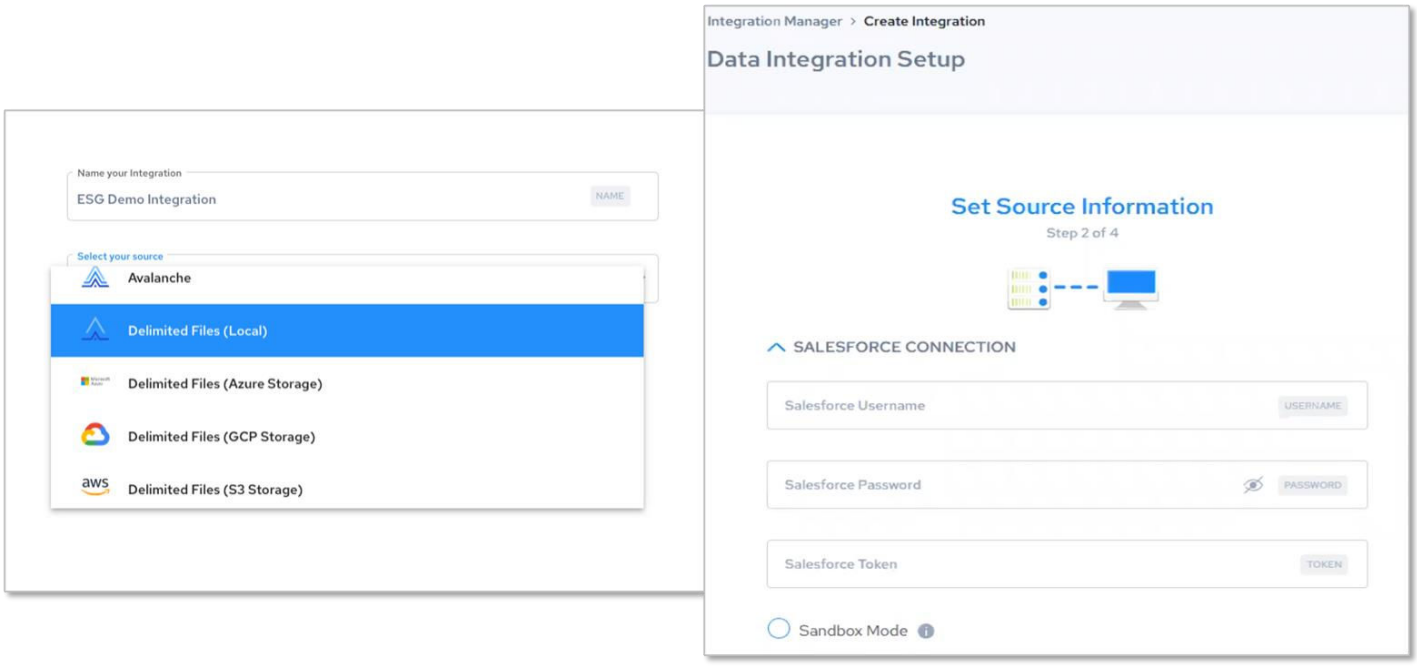
Figure 7. Automatically Mapping Source Data with Actian Avalanche Data Warehouse



Source: ESG, a division of TechTarget, Inc.

ESG navigated to the ***Integrations*** page and used a drop-down menu to select a data source (see Figure 8). After selecting Salesforce from the menu, another window prompted us to enter relevant information to access this data source. We noted that we would be using existing information such as username and password, which emphasized how the Avalanche platform has been designed to integrate third-party tools commonly used for data management and analytics.

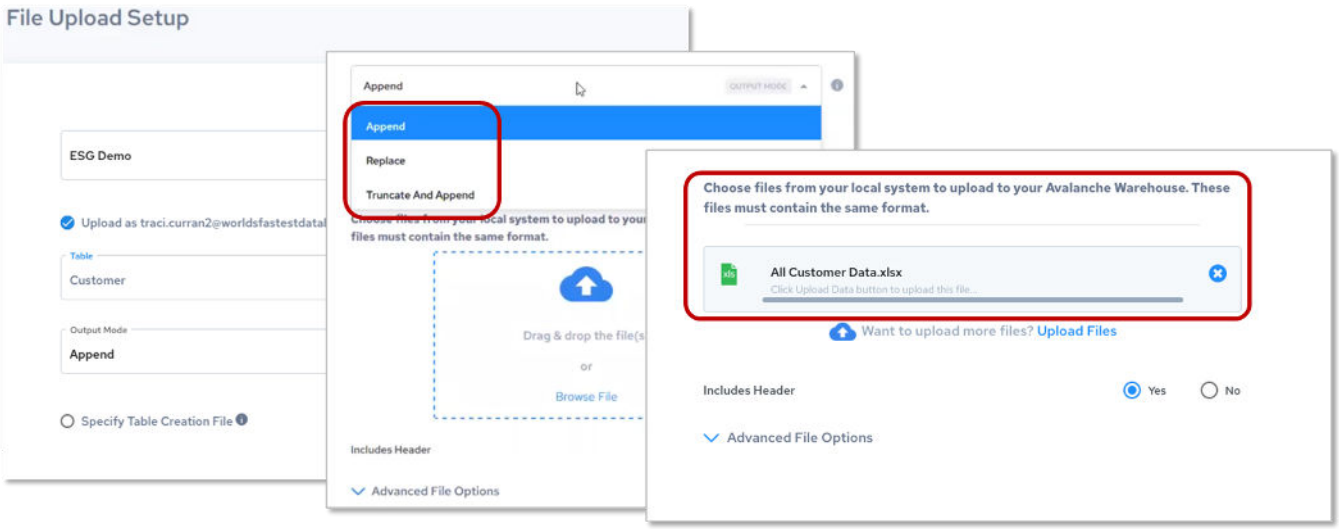
Figure 8. Uploading Select Data into Actian Avalanche Data Warehouse



Source: ESG, a division of TechTarget, Inc.

ESG also reviewed how to upload data from files stored locally on a desktop. For our "*ESG Demo*" data warehouse, we uploaded a file to populate the "*Customer*" table in the "*ESG Demo*" data warehouse (see Figure 9). After selecting how to upload the file—Append, Replace, or Truncate and Append—we chose "All Customer Data.xlsx."

Figure 9. Uploading Select Data into Actian Avalanche Data Warehouse

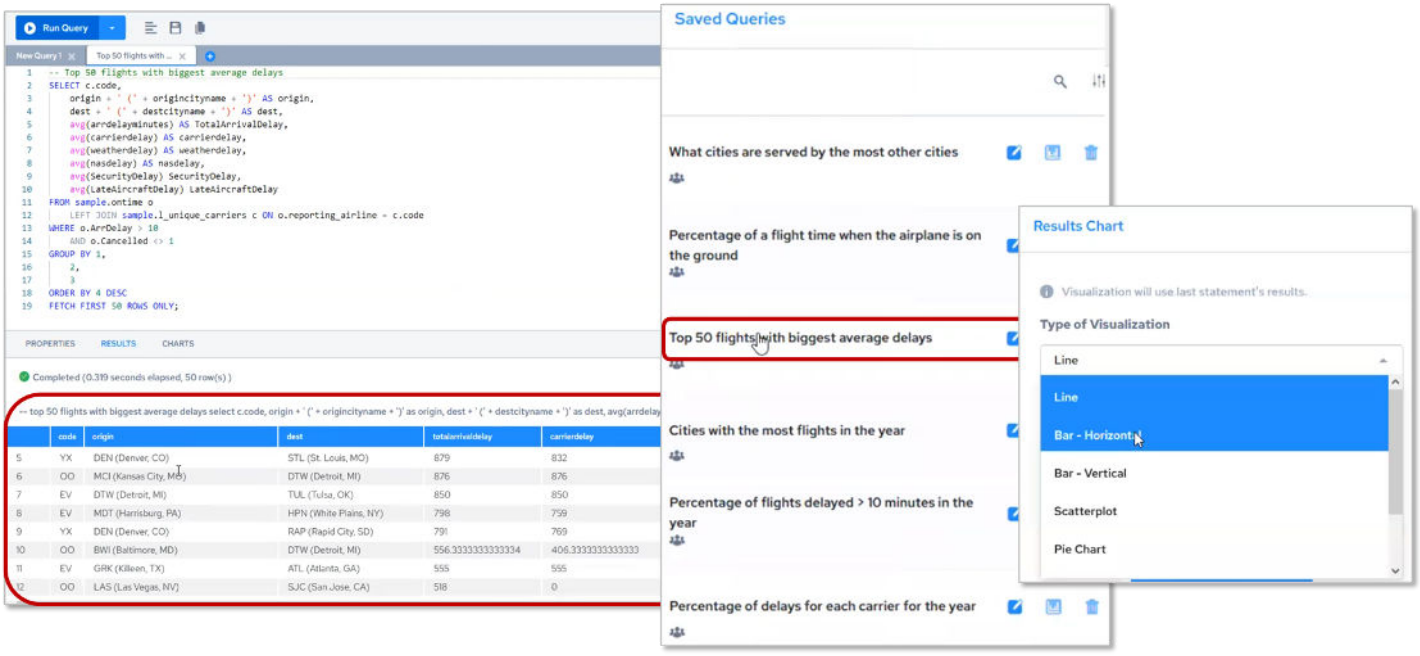


Source: ESG, a division of TechTarget, Inc.

ESG specifically noted how easy it was to populate data warehouses with the Avalanche platform. Because organizations must be able to minimize time to insight, integrating and preparing data from multiple sources must be done as quickly as possible. Spending time developing the proper ETL tools and processes that work with a wide variety of data sources only delays that time to insight. In addition, providing access to data from multiple sources also helps to improve the quality of business insights, and, subsequently, business decisions.

We finally observed the platform's data analysis capabilities and saw that it was not necessary for end-users to learn new tools (see Figure 10). Using previously developed queries (written in SQL), we queried the top 50 flights with the largest average delays. Charting options were also available to facilitate data visualization.

Figure 10. Running Queries in Actian Avalanche Cloud Data Platform



Source: ESG, a division of TechTarget, Inc.

Why This Matters

Providing access to the necessary data, especially across multiple data repositories, must be done quickly so that organizations can gain the necessary insights to meet evolving business needs. The quality of those insights is only enhanced when all relevant data is available. Yet, integrating all that data requires time and effort that prevents such insights from being obtained and used as quickly as possible.

ESG validated that Actian's Avalanche Cloud Data Platform enables organizations to gather and integrate multiple data sources, located on-premises or in the public cloud, without requiring end-users to abandon the tools they already use. We observed how easily organizations can connect with third-party applications and data repositories using native integrations and upload select data from both on-premises and cloud-based storage. The result is that end-users can maximize the quality of business insights they can gain.

Optimal Price-performance

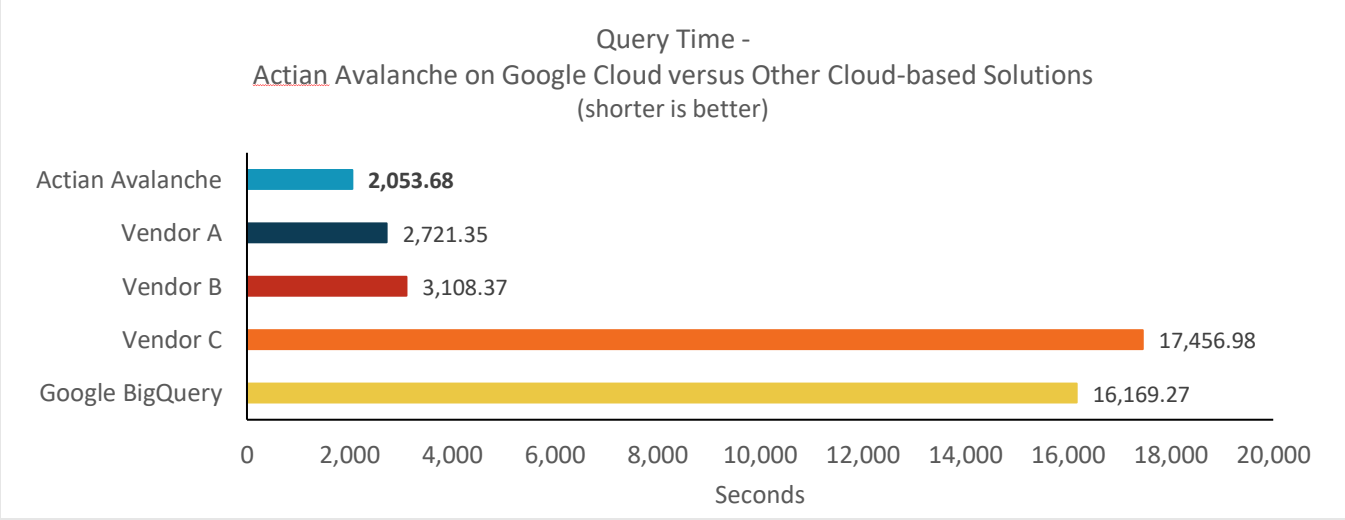
The myriad of cloud-based solutions for data warehousing and analytics presents a wide variety of cost profiles and varying levels of performance. Ideally, organizations are looking to maximize performance with minimal cost. Leveraging Google Cloud's architecture helps the Actian Avalanche Cloud Data Platform to achieve higher performance at lower price points. It is worth mentioning that Kubernetes is beneficial to Actian from both a management/engineering velocity standpoint and for the end-user, where operations such as adding or removing cloud resources can be much faster. Actian reports approximately 50% faster cloud provisioning time when using Kubernetes.

ESG Testing

ESG audited results of decision support tests consisting of an analytics workload with a combination of high data ingest and query processing similar to those performed against the TPC-H benchmark.[4] While this was not an official TPC-H benchmark, tests were designed using the setup, standards, and configurations recommended by the benchmark. Using a 30TB dataset, tests simulated queries run by five concurrent users of cloud-based data warehouses. Ad-hoc queries and generated reports associated with business intelligence were simulated. Additionally, query times using Google BigQuery were obtained since BigQuery is native to Google Cloud and is associated with big data analytics.

We first examined the query times, which measured the time for all five concurrent users to complete their queries. For each solution, the run with the longest query time was chosen for comparison. Results revealed that the Avalanche platform achieved the lowest query time (see Figure 11). Specifically, the query time achieved by the Avalanche platform was 1.3x faster than Vendor A, 1.5x faster than Vendor B, and 8.5x faster than Vendor C. Compared to Google BigQuery, Actian Avalanche was 7.9x faster.

Figure 11. Comparison of Query Times

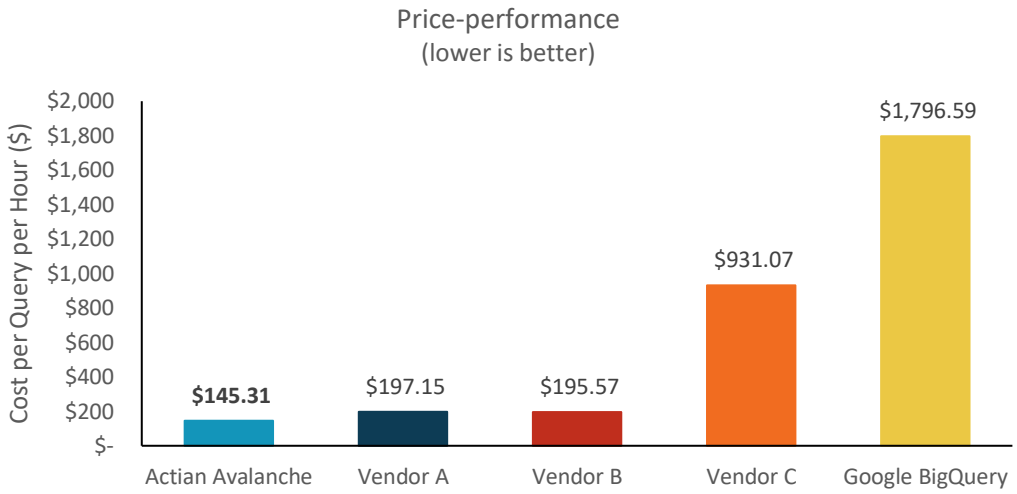


Source: ESG, a division of TechTarget, Inc.

Costs for configuring and running similar testbeds on each cloud-based solution were estimated using publicly available information. Using the obtained query times, cost per query per hour was then calculated. Results revealed that the Avalanche platform achieved the lowest cost per query per hour (see Figure 12). Specifically, the cost achieved by the Avalanche platform was 26% lower than Vendor A and Vendor B, and 84% lower than Vendor C. Compared to Google BigQuery, Actian's Avalanche Cloud Data Platform cost was 92% lower.

[4] For more information about the TPC-H benchmark, please go to https://research.gigaom.com/report/high-performance-cloud-data-warehouse-performance-testing/.

Figure 12. Comparison of Price-performance



Source: ESG, a division of TechTarget, Inc.

ESG noted that the comparison of the Avalanche platform and Google BigQuery addresses those who question choosing one platform over the other, despite both Avalanche and BigQuery using Google Cloud as the underlying cloud architecture. Some may argue that using Google's native solution, Google BigQuery, is ideal for analytics of workloads requiring petabytes of data. This makes sense since with BigQuery, the more data that is used, the more cost-effective the solution. On the other hand, Actian's Avalanche Cloud Data Platform is designed to run analytics for organizations accustomed to working with terabytes of data at any given time. However, these two solutions are not mutually exclusive and can be used together to provide flexibility for analytics teams, giving them the option to choose the best environment for their unique use cases.

Why This Matters

While ensuring overall performance of a cloud-based data warehouse is critical when considering how quickly business insights can be obtained, the costs for running queries vary widely amongst the available options. Minimizing price-performance thus remains a key criterion in choosing an appropriate solution.

After auditing results obtained using tests modeled against the TPC-H benchmark, ESG validated that Actian Avalanche on Google Cloud can achieve the lowest cost per query per hour for five concurrent users compared with other well-known cloud-based data warehouses. When comparing the Avalanche platform with Google BigQuery, we found that using the Avalanche platform for queries against terabytes of data is preferable to using Google data analytics solution, as it is more appropriate for analytics workloads requiring petabytes of data.

The Bigger Truth

Teams across organizations and industries that analyze and seek insights from data are often diverse, crossing multiple disciplines. While they leverage many of the same data sources, they have disparate needs, are trying to answer different questions, and are often using different tools.

Gaining quality insights requires the ability to draw on data from every corner of an organization and make it available to everyone who needs it—from internal stakeholders to partners and customers outside the company. Data must be able to move in any direction as it is enriched, processed, and merged with other data—with a goal of uncovering new opportunities for products, services, and partnerships.

ESG validated that the Actian Avalanche data platform:

- Makes it easy for users to configure and dynamically scale data warehouses, removing the need to manually choose and size compute instances.
- Enables organizations to gather and integrate multiple data sources, located on-premises or in the public cloud, without requiring end-users to abandon the tools they already use.
- Makes it easy for organizations to connect with third-party applications and data repositories, using native integrations to leverage data from both on-premises and cloud-based storage.
- Achieves the best price-performance—up to 7.9x faster with up to 92% lower cost—for five concurrent users compared with other well-known cloud-based data warehouses.

Additionally, analysis of testing confirmed Actian's claim that Avalanche can provide exceptional performance for operational workloads, with Actian stating that most operational datasets are under 300TB.

The results that are presented in this document are based on testing in a controlled environment. Due to the many variables in each production environment, it is important to perform planning and testing in your own environment to validate the viability and efficacy of any solution.

If your organization is looking to extract actionable and accurate insights at the best price-performance to drive a wide range of use cases—from risk assessment and analysis in financial services to network monitoring in telecommunications to market basket analysis for online shoppers, the Avalanche Cloud Data Platform on GCP is worth your serious consideration.