Faster Together: Actian Avalanche on Google Cloud Platform

Gain a 6X Price-Performance Advantage over Snowflake on AWS

Actian Avalanche on Google Cloud Platform (GCP) delivers six times the price-performance of Snowflake on AWS, enabling organizations of all sizes to run ad hoc queries on billions of data records—with sub-second response times. You can affordably fire up additional real-world, operational BI, and analytics workloads while enabling concurrent access to more users than ever. In short, Actian Avalanche on GCP provides the price-performance advantages you need to translate “data-driven” into tangible real-world outcomes.

Whitepaper
April 2021

The Challenge: Expanding Access to Affordable Real-Time Decision-Making Capabilities

New York Times columnist Thomas Friedman wrote a best-selling book, engagingly entitled, *The World is Flat*. It describes a world of immersive globalization, pandemics, hyperconnectivity, and financial interconnectivity, a world in which no business is too small or too large to fail. Competition is fierce. Businesses must be agile, and agility depends on the intelligence informing an organization's business processes. How well do your business analysists and critical decision-makers understand every aspect of your business, employees, and customers? Can they derive actionable insights from data? Equally important, can they *respond* at the speed of business?

Strategic capabilities based on real-time decision-making are key, and at the core of all of them is the ability to run BI and advanced analytics. Whether strategic capabilities involve overall risk assessment and management or a modification of specific business processes—in areas as diverse as demand planning and forecasting, customer experience analysis, security and fraud analysis, and more—your ability to act depends on your ability to extract and make meaning of data. And not just any data: it must be the *right* data, from the *right* systems, including those at diverse points of engagement such as mobile handsets, the IoT, and far-flung intelligent equipment.

Few organizations are well-positioned to enable the diverse group of experts who need this data to draw those deep insights quickly, easily, or affordably. On-premises infrastructures like those based on Hadoop are neither fast, readily accessible by non-IT personnel, nor cost-effective (particularly when accommodating large volumes of data). Consequently, users in most organizations do not rely exclusively on their on-prem infrastructures. Online services—including Salesforce, Marketo, and many others—play an ever-increasing role in both day-to-day operations and critical business decision-making. But from the perspective of insight discovery, these external solutions create further complications. Consolidating and normalizing the data from external and internal, on-prem and cloud-based, structured and unstructured sources requires an enterprise data warehouse (EDW) and involves complicated and time-consuming data ingestion and data transformation processes understood only by those individuals initiated into the mysteries of the IT department. For those seeking insight, even gaining *access* to the data in an EDW is not easy. That too requires the intercession of IT at too many points, resulting in both time-to-insight delays and higher costs.

Actian Avalanche

Best Price Performance

- Pay only for what you use
- Transparent and predictable costs as you add concurrent users, data, ad hoc analytics volume, and complexity
- Scale compute and storage resources independently—up and down.

Fully Managed Service for Analysts, Engineers, Scientist and Power-Users

- Turnkey offering on Google Marketplace with by-job billing and management
- IT involvement needed only to the extent your corporate guidelines require
- Built-in data connectors, with menu-driven support and out-of-the-box integration
- Meets SOC-II level enterprise security

Google Cloud

Design Optimized for Google Cloud

- Designed from the ground up for Google's cloud ecosystem
- Containerized to take advantage of Google Kubernetes Engine (GKE) management
- Fully integrated with Looker for real-time visualization and reporting

Nor have organizations that have migrated wholly to the cloud fared much better in overcoming these challenges. The initial generations of cloud data warehouses merely forklifted existing shortcomings into the cloud. Early appliance-centric offerings from companies like Teradata, Netezza and Oracle Exadata were unable to take advantage of the elasticity of the cloud to provide pay-as-you-go, dynamically-scale-up-and-down offerings. Newer solutions, such as Amazon Redshift, Microsoft Synapse, and Snowflake, provide scalable, pay-as-you-go offerings that support real-time analytics involving massive volumes of data. However, all these are costly solutions that continue to require the intermediation of IT to consolidate and prepare disparate datasets and connect users to those consolidated datasets.

In short, affordable direct access to the data that users and business processes really need, in a manner friendly enough for non-IT users to navigate without IT intervention, has long remained an unchecked box—until now.

The Solution: Actian Avalanche on GCP

Actian Avalanche and GCP work together at foundational levels to deliver an affordable, dynamically scalable cloud-based EDW solution designed to blend maximum access to insight with maximum ease of use. Suitable for organizations of any size and offered as a managed service through Google Marketplace, Avalanche takes full advantage of the elasticity of GCP. Compute and storage resources are managed independently, so Avalanche can automatically add more CPU cores and RAM to deliver the performance you need; it can also automatically remove unused CPU and RAM capacity if those resources are idle. The same thing is true with storage: You always have the cloud storage resources your datasets require, and the underlying Google infrastructure will add and remove storage as needed.

As for the cost of such responsiveness? Beyond a minimal monthly service fee, subscribers are charged only for the resources used. You never need to overbuild to support occasional spikes in demand because Avalanche and GCP automatically scale to accommodate those spikes while still delivering the performance you expect. When the spike subsides, GCP scales your resources back to baseline and you only pay for the added resources while they were in use during the spike.

A Peek at Peak Performance

Consider performance, first. While industry-standard benchmark tests such as the TPC Benchmark™ H (TPC-H) are somewhat artificial—compared to real-world use cases which may be more or less dynamic—the value of such benchmarks lies in their ability to showcase differences in performance against the same datasets and queries. TPC-H involves a suite of 22 queries run against a 30TB dataset consisting of billions of records in separate tables representing customers, products, orders, parts, suppliers, and more.

Independent auditors put Avalanche to the test using TPC-H and, as shown in Figure 1~~Figure 1~~, compared performance measurements to those associated with Snowflake running on AWS.

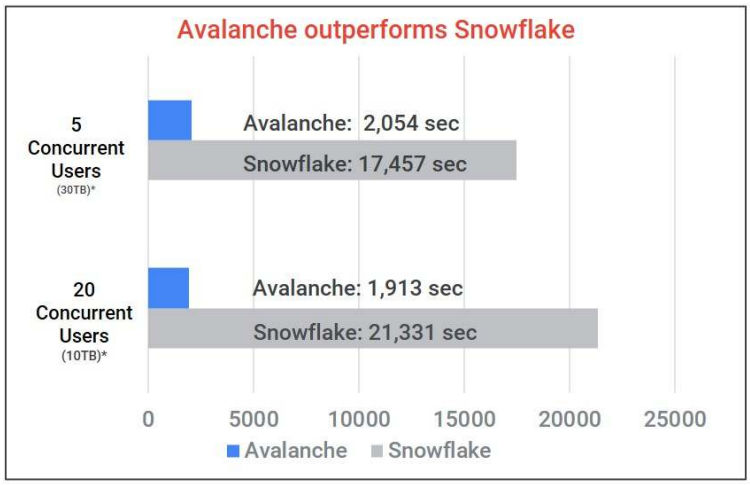

Figure 1: TPC-H benchmark comparisons (Avalanche vs. Snowflake)

In benchmarks simulating five concurrent users interacting with a 30TB dataset, Avalanche delivered query responses 8.5 times faster than Snowflake running on AWS (note: Avalanche was run on AWS to create an apples-to-apples comparison; it runs 20% faster on Google Cloud than AWS). In simulations involving 20 concurrent users and a 10TB dataset, Avalanche performed more than 11 times faster than Snowflake on AWS.

Priced to Perform

Historically, the cost of an EDW capable of performing with the speed of Avalanche on GCP—particularly on a 30TB dataset—was prohibitively high. The up-front CAPEX investment alone was significant—particularly if the organization over-configured in anticipation of demand spikes—as were the ongoing OPEX for monitoring, maintenance, and IT support for end-users.

The cloud, though, offers a way around many of these costs. Organizations no longer need to overbuild an on-prem infrastructure when all the power required is available in the cloud. Similarly, OPEX overhead is reduced if the organization no longer needs skilled personnel to manage an on-prem data center; such tasks are performed by personnel associated with the cloud service itself.

What is needed is an EDW architecture designed to take maximum advantage of the elasticity of the cloud. As previously noted, Avalanche takes advantage of the separate compute and storage infrastructures informing GCP. It can dynamically add and remove compute and storage resources—in real time—to deliver against your performance expectations even when there are spikes or slowdowns in demand. And while customers will pay more as more resources

"In our testing, Avalanche query response times on the 30TB TPC-H data set were overall 8.5 times faster than Snowflake in a test of 5 concurrent users. Examining price-performance, Avalanche ran the 5 concurrent user TPC-H queries roughly 6.4 times more cost effectively than Snowflake, as measured in cost per query per hour."

William McKnight and Jake Dolezal, GigaOm & MCG TPC-H Benchmark Study, October 2020

are called into service, they will always pay only for the resources used. Moreover, the resources used will always be commensurate with the performance demanded.

How does this play out in terms of price-performance? The same auditors who compared the TPC-H performance of Avalanche against Snowflake on AWS calculated the price-performance difference between the systems, based on a comparison of cost per query per hour. Their findings showed that Avalanche is 6.4 times more cost-effective than Snowflake running on AWS.

Figure 2: Actian Avalanche is 6.4 times more cost-effective than Snowflake on AWS

Better Insights Lead to Better Outcomes

Benchmarks can be impressive abstractions. The real questions an organization needs to ask boil down to these: How much will a given insight cost? Can we gain that insight ahead of the speed of the business itself? In terms of the actual price-performance of a cloud data warehouse, those questions play out in different ways in different scenarios. Consider:

- An AI-driven recommendation engine needs to retrieve all customer-relevant options in order to formulate an appropriate recommendation in a real-time customer engagement situation. In many cases, businesses face a tradeoff: Either spend more than they can afford to meet the time constraints, or receive stale or delayed insights to avoid cost overruns. In terms of price performance, will a delay in the delivery of recommendation options mean the loss of an opportunity to cross-sell or up-sell, or will the cost of scaling the compute resources to increase query throughput push the aggregate cost of the sale outside the budget for that company or department?

- A healthcare insurance underwriting engine needs to pull a wide range of data—including credit ratings, demographic information, predictive medical information, and more—from disparate sources to determine whether, and at what price, to offer an applicant an insurance policy. Can your company accurately estimate how many concurrent users will interact with your EDW during open enrollment season? Is your EDW provisioned to ensure uninterrupted performance during those periods? Can you scale up rapidly, transparently, and predictably if business demands require it?

- An unfamiliar customer is depositing more than $10,000 dollars into a bank account. Can you run a query of their bank records and the anti-money laundering watch list quickly enough to avoid impacting

customer service (or offending the customer if they are acting in good faith) or to alert the authorities if their actions are illicit? What is the cost of gaining such insight in a timely manner? What is the cost of *not* being able to gain it in a timely manner?

In all these cases—and countless others—the price-performance of traditional EDWs or even the first generation of cloud EDWs can preclude access to critical insights. With the price-performance advantages delivered by Actian Avalanche on GCP, rapid access to critical insights quickly translates into better business outcomes.

Make "Data-Driven" the Reality for All Organizations

Every organization *claims* to be data-driven, but the degree to which this is true varies. It's not that most organizations are lying; it's more that the insights driving decisions are based on a highly fragmented and incomplete analysis of the data they can access. To claim that the organization is data-driven is like claiming to have read *Moby Dick* because you have read the first chapter.

For most businesses, the problem lies in gaining a unified view of the data they have. Some of it resides in siloed SaaS and legacy solutions. Some of it is highly structured; some of it is unstructured; some of it may even be semi-structured. Some of it is persistent; some of it is transient and always in motion. Yet when business analysts, data engineers, and non-IT power users are striving to make well-informed decisions, *some* parts of *all* of that data is important. All of it needs to be accessible. Whether one is running complex BI queries, performing advanced analytics, or training a machine learning engine, the ability for non-IT personnel to access and interact with the data they need—without the intermediation of IT—is crucial to the acquisition of deeper and stronger insights that can truly drive business.

Actian Avalanche on GCP provides a flexible, powerful, and demonstrably more cost-effective way to unify all your data and derive from that data the insights your experts are seeking. Providing more than 6x the price-performance benefits that Snowflake on AWS can offer, Actian Avalanche on GCP ensures that your EDW is always right-sized, can scale up or down automatically in response to user and enterprise demands, and can provide easy access to virtually any dataset your data scientists, analysts, and power users might want to factor into their models and explorations.

Test drive Actian Avalanche today by selecting it in Google Marketplace. For more information about Actian Avalanche on GCP, visit our website at actian.com/partners/google-cloud.