

INSIGHT PAPER

Accelerating Analytics with Alteryx and Snowflake

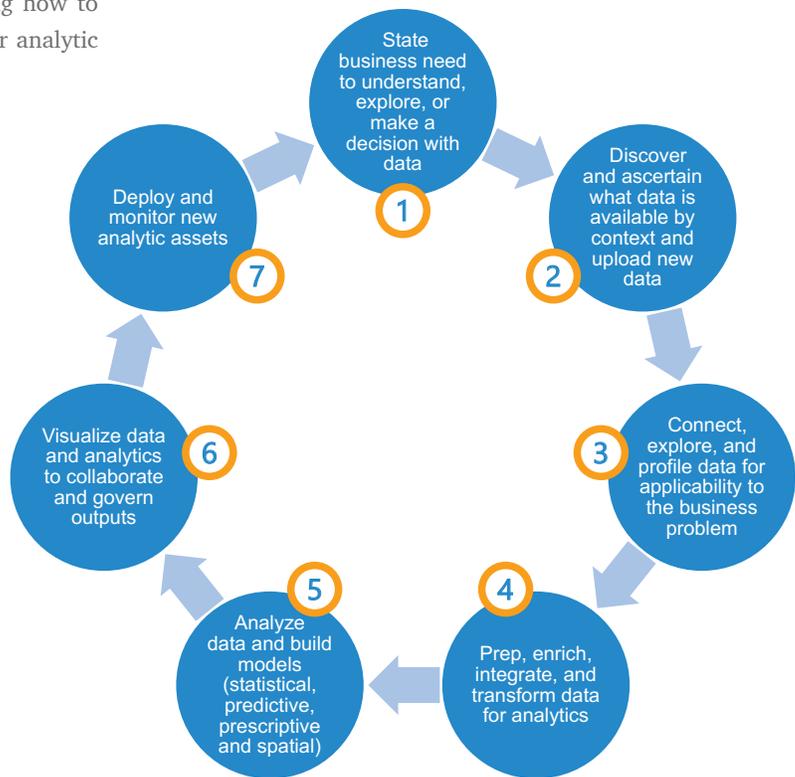
Analytic leaders with the responsibility of unlocking the business value of enterprise data assets continually strive to empower people throughout the organization to work with data in more powerful ways. To do so, these leaders must have a firm understanding of how people interact with the technologies and processes supporting the full lifecycle of modern analytics. With this knowledge, they are able to guide the expansion of an enterprise data environment with technology and approaches that remove the barriers to self-service data analytics and insights.

Cloud-based services combined with powerful self-service analytics platforms, such as Alteryx with Snowflake, facilitate many of the functional characteristics that these leaders need to help people thrive with modern analytics. Creating the environment to support a data-centric culture takes time, but clearly recognizing how to relate specific technologies to empower people and their analytic process will naturally propel enterprise adoption.

The **Modern Analytic Lifecycle** based on Radiant Advisors research breaks down the cognitive steps that people go through to solve problems with data by incorporating people, processes, and technology into a cyclical workflow to support data centricity and competency. Analytic leaders seeking to make better decisions faster and with more confidence must focus on removing the friction or delays within each lifecycle step while increasing the context and understanding of the data and work being performed. Each of the seven steps of the lifecycle surfaces four themes that a modern data platform environment must provide: agility, self-sufficiency, high performance, and scalability.

From this perspective, the **modern data platform** needs to fulfill an architecture requirement to facilitate self-service data analytics for anyone to access and work with data easily and intuitively. As part of this architectural design, many organizations incorporate a public cloud platform, such as Amazon Web Services (AWS) or Microsoft Azure, in their data strategy to take advantage of affordably scalable services and on-demand resources with usage-based billing. This allows organizations to keep up with the growing demands for data and analytics without the burden or delay of internal infrastructure procurement, provisioning, and management.

Figure 1: The Modern Analytics Lifecycle



Accelerating Analytics with Alteryx and Snowflake

INSIGHT PAPER

Self-service data science and analytics platforms, such as Alteryx, are a primary component in a modern data platform architecture to empower anyone to work directly with data and develop the analytics they need. Self-service data analytics provide an agile and intuitive way for people to find, connect, transform, and share analytics throughout departments, divisions, and the company overall. To help business analysts and data scientists get the most from cloud-based self-service analytics, Snowflake – a cloud-native data warehouse as a service – can address workload roles and scalable high-performance requirements that user activity and scheduled jobs require to work with both structured and semi-structured data.

The Business Value of Alteryx Combined with Snowflake

The modern data platform allows people to move quickly and confidently through the analytics lifecycle with improved agility, self-sufficiency, responsiveness, and scalability. Some steps in the lifecycle are compounded by a highly iterative, user-driven process that is repeated as many times as necessary before moving to the next step. For example, exploring and profiling a prospective data set may require dozens of view refreshes, re-sorting, filtering, and joins to determine whether it fits the business need before moving to explore another data set. Alteryx provides the visually intuitive user interface while Snowflake provides all of the data quickly to avoid interruption to a data worker's thought process while working.

Alteryx combined with Snowflake increases overall data availability and freshness. For data ingestion, many companies will consolidate their miscellaneous and legacy data ingestion scripts or jobs into Alteryx Server workflows to be centrally scheduled and monitored for reliably delivering data to users. For ingesting data, the Alteryx high-performance bulk data loader tool is specifically made for Snowflake and delivers faster load times for large output data sets versus traditional database loads and inserts. Alteryx workflows that use Snowflake as a data source to integrate data reduce execution times and make the data available quicker in Snowflake data targets, thus resulting in fresher data that can be accessed sooner by data visualizations, reporting, or data science

tools. Snowflake extends similar data availability benefits to any of its computing clusters, including the enterprise data lake, data warehouse, data marts, and analytic databases.

The combination of Alteryx and Snowflake improves overall data analytics performance for data transformations and related queries by leveraging computing power scalability in the cloud and Alteryx tools for Snowflake. Both Alteryx and Snowflake benefit from cloud scalability for easily adding additional Alteryx Server worker nodes or increasing Snowflake computing cluster sizes when additional computing resources are required for improving performance. The Alteryx In-Database tools execute workflows in Snowflake for faster performance of data manipulation and data science calculations that avoids data transfers in and out of the database. These in-database manipulations also benefit users developing local workflows involving large data sets with Alteryx Designer by avoiding costly data transfers out of the cloud to the desktop. Further, executing Alteryx workflows in a dedicated Snowflake computing cluster ensures that its workload does not impact other workloads for individual user query response times and busy user concurrency by application or group. This also simplifies analytic workload management.

Alteryx with Snowflake also provides scalability and self-sufficiency to meet current and future business needs without hindering adoption of self-service data analytics over time. Alteryx Server components easily scale with more cloud virtual machines or with cloud auto-scaling configurations that can grow and shrink compute resources based on demand. Rather than estimating the future size requirements of on-premises servers to grow into, cloud platforms allow business groups and IT to add resources within minutes and remove resources if or when needed. Snowflake's computing clusters are monitored and can be scaled according to their workload needs. Because Snowflake stores its data in a highly efficient compressed format, its data capacity can be considered limitless. In the case of a self-service data analytics environment, new computing clusters can be provisioned quickly, or a single large analytic sandbox can be provisioned and sized to handle the ad hoc data discovery workload.

Accelerating Analytics with Alteryx and Snowflake

INSIGHT PAPER

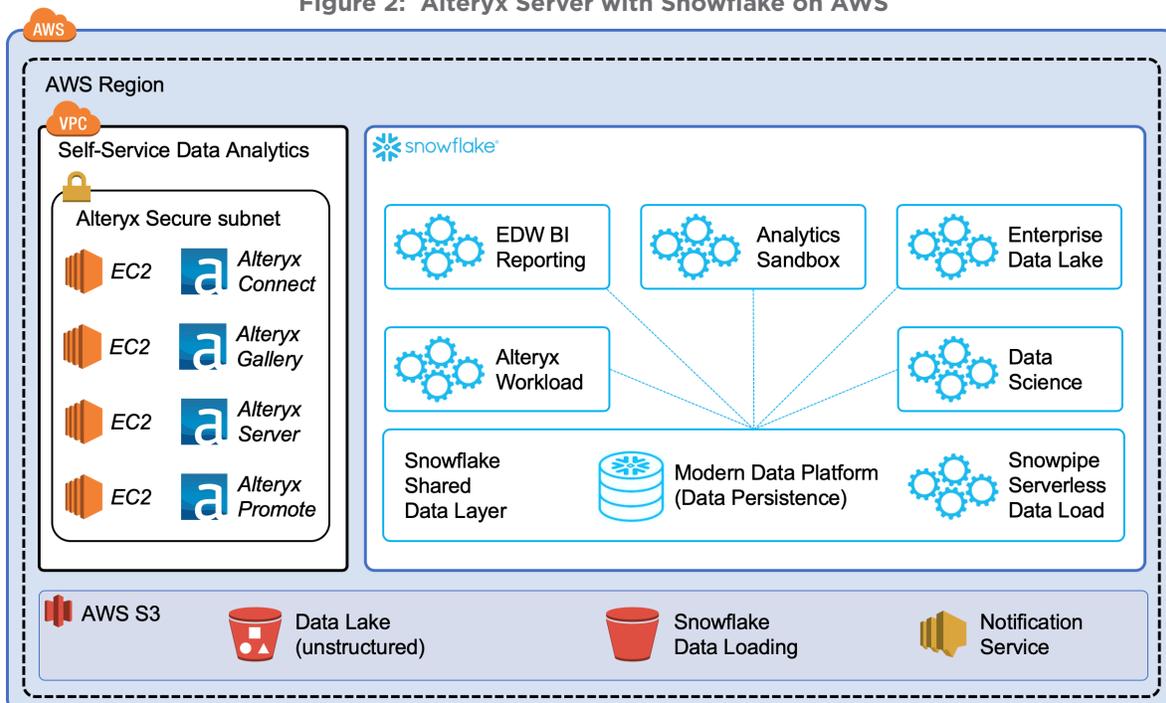
Digging In: What Analytic Leaders Need to Know

Alteryx Server in the cloud simplifies the scalability and high availability requirements for all of the workflows that business analysts, data engineers, and data scientists will create and need to run. When workflows become an integral part of business operations, they will need to be monitored and architected in such a way to ensure reliability – which is provided more easily with cloud platforms. The key components of the Alteryx Server (Controller, Worker, Gallery, and Database) can work as a single scalable system and be installed and running on as many cloud virtual machines as needed. To further optimize the Alteryx Server, Radiant Advisors recommends analyzing workflow execution and modifying the deployment architecture to account for its components individually, thus taking advantage of usage-based billing for only the virtual machines in use.

Business continuity with Alteryx can be ensured when Alteryx products and components are architected for deployment to public clouds that transparently provide high availability and disaster recovery features with Infrastructure as a service that leverages multiple cloud data center regions and availability zones. This will ensure that Alteryx Server, Alteryx Connect, and Alteryx Promote have the high availability and resiliency needed for production workloads from the individual users to the enterprise scale. Whether it's installed as a single server or cluster of servers, cloud platforms simplify how to scale reliably versus on-premises data center installations that must also have disaster recovery plans and facilities.

Snowflake is a data warehouse as a service with a cloud-native architecture optimized for both AWS and Azure. Snowflake provides many of the database roles and instances necessary

Figure 2: Alteryx Server with Snowflake on AWS



Accelerating Analytics with Alteryx and Snowflake

INSIGHT PAPER

in a modern data platform for delivering high scalability, high-performance analytics, and high user concurrency. Snowflake follows the cloud architecture design pattern of decoupling data storage from computing resources in order to scale them independently and on-demand for optimal performance. Therefore, Snowflake is a usage-based computing service that stores all of the data together and allows for easily creating and managing clusters of cloud virtual machines for groups of users or workloads to access.

The Snowflake virtual warehouse is a term to describe a cluster of many cloud virtual machines that are assigned to execute database operations for a specified group of users or database workload. This concept allows Snowflake computing resources to access shared data as virtual warehouses of different sizes as needed to support different workloads, such as the data lake workload, an enterprise data warehouse workload, and an Alteryx processing workload. These virtual warehouses isolate their activity from each other and are tailored to meet their specific workload with the right number of vCPUs and memory. The ease of administration in creating and managing virtual warehouses from abundant cloud resources makes it easier to meet user needs than when working with traditional on-premises data centers.

The enterprise data lake provides a repository of all enterprise data assets to be leveraged as part of the self-service data analytics environment by anyone for business intelligence, self-service data analytics, and data science initiatives. Cloud-based data lakes are typically characterized by the public cloud's object storage service, such as AWS S3 or Azure Blob Storage, and early data lakes may have been deployed on Hadoop clusters with the Hadoop Distributed File System (HDFS). Object stores facilitate the unique ability of data lakes to provide low-cost, scalable, and resilient storage for any data format, including structured data files, document files, image files, and video. That said, companies may choose to deploy the structured and semi-structured portion of the data lake in relational databases due to familiarity, SQL access, security, and the fact that the majority of source data needed for

analysis is from structured data sources or semi-structured files. In these instances, Snowflake's columnar compressed data storage, semi-structured data support, and SQL database engine provide familiar tools for users and higher performance versus alternatives, such as traditional data warehouses.

Bringing it Together

The strengths of Alteryx and Snowflake combine to sustain and optimize the modern analytics lifecycle while improving user experience and establishing the foundation for enterprise data analytics scalability. In doing so, their combination within a modern data platform architected for analytics is able to increase user adoption and advance data-centric company cultures.

For more information and detailed reference architecture diagrams, please read the corresponding technical white paper, "Modern Analytics Architecture: Alteryx, Snowflake Computing, Amazon Web Services, and Microsoft Azure," available at www.alteryx.com.

Brought to you by:



Revolutionizing business through data science and analytics, Alteryx delivers a modern, end-to-end analytics platform empowering analysts and data scientists to deliver game-changing insights like never before.

To learn more, visit www.alteryx.com



Radiant Advisors is a leading strategic research and advisory firm that delivers innovative, cutting-edge research and thought-leadership to transform today's organizations into tomorrow's data-driven industry leaders.

To learn more, visit www.radiantadvisors.com