Actian Avalanche Hybrid Federated Query

Let your data live where it naturally resides

Key Features

Enables real-time analytics by sourcing data directly from the system of record

Lowers overhead and improves query performance by minimizing data movement

Reduces cost and complexity by decreasing the need for creating and managing ETL processes

Despite the hype surrounding cloud computing, most organizations still have a large portion of their business-critical data on-premises and the rest spread across multiple clouds. This has led to the creation of more data silos and reduced the ability to analyze entire datasets for extracting insights.

Hybrid data environments are likely to continue to persist for several reasons:

- Organizations want to retain complete control over sensitive datasets, particularly those that are subject to regulatory compliance requirements
- Data producing apps are still on-premise (e.g. legacy ERP), which makes it logical to have the corresponding data stores on-premise as well
- Many organizations have made a significant investment in on-premise infrastructure and want to shift to the cloud gradually

Hence, forward-thinking organizations are incorporating a hybrid strategy that can query large sets of data across on-premise and across multiple clouds in order to extract meaningful insights.

Not All Data Can or Should Be Moved to a Centralized Repository

IT leaders are increasingly accepting the necessity of distributed queries for these hybrid data contexts. Unfortunately, executing queries that span cloud and on-premise data warehouses by moving data to a single repository poses numerous challenges. For one, it may be too costly to create a separate copy of the data in a centralized location. Secondly, some data may be subject to specific regulatory requirements that dictate how it is collected, stored and processed.

Enterprises should instead look for ways to access data as if it were in a virtual data warehouse, going to the data where it naturally resides rather than copying it everywhere in anticipation of its being needed later. However, many data warehouse solutions have limited deployment options. Appliance-based solutions, for example, cannot be deployed to the public cloud. Virtualized versions of these solutions may not perform optimally on commodity hardware in the cloud.

Introducing Actian Avalanche Hybrid Cloud Data Warehouse

Actian Avalanche is a fully managed hybrid cloud data warehouse service that is purpose-built to address the demands of today's distributed hybrid data environments. It can seamlessly be deployed on-premise as well as on AWS, Azure and Google Cloud. All deployments, regardless of location, are compatible so there is only one technology stack for you to learn, deploy, and manage. It provides the perfect solution for on-premise datasets that cannot move to the cloud due to compliance requirements or other business restrictions.
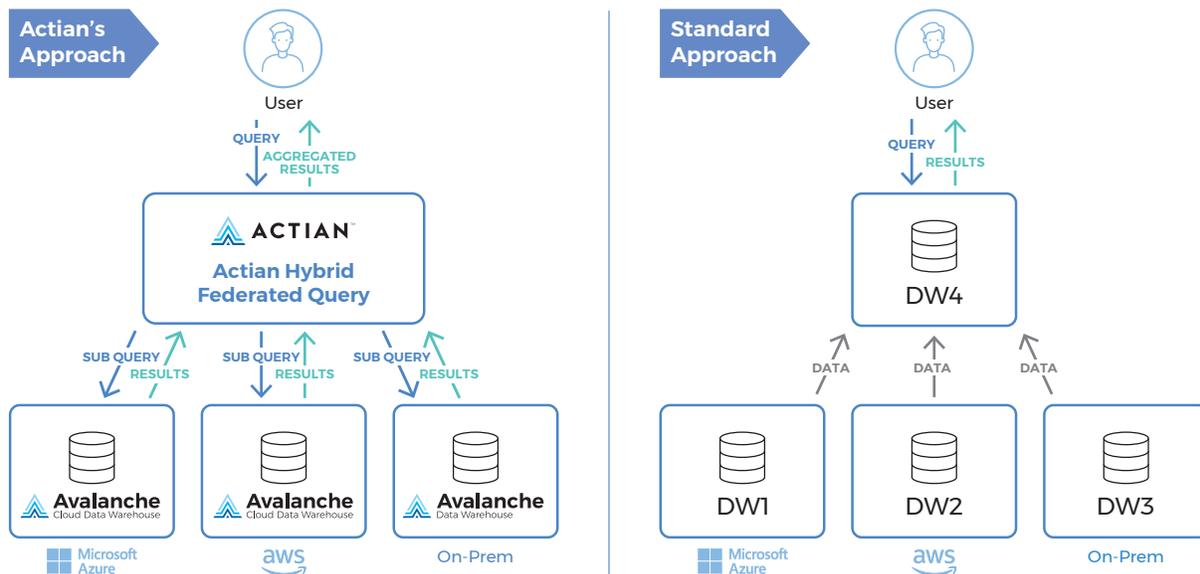
Figure 1. With Actian Avalanche, you don't need to move all your raw data to another repository. Compute is moved to where data resides.

Actian's Hybrid Federated Query Reduces Complexity and Lowers Cost

Actian Avalanche hybrid cloud data warehouse features intelligent federated query capability that can query on-premise and cloud data without the need for moving the base data to the cloud.

Avalanche uses integrated Spark technology to access external data sources, registers the data from those sources as external tables within Avalanche, and treats that data like its local Avalanche data. Alternative solutions often require extensive custom coding and other integration products to access and ingest external data. With Avalanche, everything you need to conduct federated queries of hybrid data sources is built into the service offering.

Advanced capabilities such as the ability to restrict portions of a dataset from being accessed, e.g. limiting the dataset to a particular zip code or account ID, can be extended to apply to the remote source so that only the data required to solve the query will be brought into Avalanche.

Process data where it resides for real-time analytics

In general, it is optimal to minimize data movement as ingress and egress from the cloud is expensive and slow. Processing and querying data where applications are producing it often produces the best results in terms of cost and performance. If a marketing campaign is being conducted in the cloud, for example, it makes sense to do the analysis in the cloud. Analyzing hundreds of on-premise ERP application logs is naturally an on-premise task.

Avalanche can seamlessly distribute and execute queries across both on-premise and cloud instances of Avalanche without using an ETL tool or forcing data movement across sources. You can use a single query to retrieve insights from data distributed across multiple sources while keeping the data in its natural location—on-premise or cloud. The query will automatically be distributed to run in the right place, and you will receive the results from wherever you initiated the query.

For example, a large retail chain may manage its vast historical sales data on Avalanche running on AWS. An Avalanche instance running on-premise in a corporate data center may contain current supply chain data, and customer sentiment could be delivered in real time via a Kafka stream. The procurement manager's BI dashboard would draw from all three data sources transparently, and queries would be run on both instances of Avalanche with no changes required.

By sourcing data directly from the system of record instead of loading data in a batch process into a single data warehouse, you will lower cost and realize faster time to insight.