Technology Insights... January 15, 2023

J.Gold Associates, LLC. Northborough, MA 01532 USA
www.jgoldassociates.com +1-508-393-5294
Research, Analysis, Strategy, Insight

Solving the Real Data Analytics Problem

A Publication For Clients of J.Gold Associates

"...With an increasing use of hybrid data locations and multi-cloud solutions, it is imperative that enterprises be able to run needed analytics on all of its dispersed data in order to stay competitive. Further, while many analytics solutions exist that are tied to specific data products (e.g., hyperscalers, dedicated SaaS data warehousing), the ability to ingest data from a wide variety of sources is lacking in most of these products. Finally, the need for near real time analysis by a wide variety of business users requires a solution to create data analysis functions using low code/no code capabilities...."

Enterprises rarely have all of their data in one location. Specialized environments like IoT and remote locations often have their own data storage facilities and may even have non-compatible data bases. The multi-location and hybrid cloud approach means that data is often scattered across many locations.

Companies have focused in the past on trying to bring all the data into a single data warehouse, but this approach fails to address the need to keep data close to the point of use, and it runs afoul of many geographical regulations that require localization of data storage. Further, keeping a variety of structured and unstructured data types available by utilizing a data lakehouse, while offering a major improvement over the traditional data warehouse, still creates difficulties for users who must deal with a wide variety of data types that may not always be compatible with available analytics tools. Being able to perform analytical functions across all data types and data locations is mission-critical if a full view of operations is required.

It's imperative that for enterprises to get maximum benefit from data analysis and create the most effective insights, there needs to be a way to merge all of the data into one compound data entity, even if only for purposes of analysis. Data ingestion therefore needs to be comprehensive, and it needs a solution that understands how to access data from a variety of different data bases, data repositories, warehouses, cloud-based storage, etc. It may even include ingesting data from a competitive analytics solution as some companies indeed have multiple solutions. Without such a solution, companies often have only a portion of their data available for analysis, providing only partial insights that may miss critical trends or potential difficulties.

Creating a solution can be a daunting task that requires a wide array of skills, knowledge about data sources, and resources beyond what IT traditionally provides. This often results in critical data not being integrated into analytical operations. Indeed, we estimate the average enterprise currently uses less than 50% of its data when running corporate analytics. What are they missing by not being able to include the other half of their data?

Many data analytics solutions consist of high end programmable functions that are reserved for data scientists and/or high end developers. This creates a bottleneck for line of business users who may wish to run an analysis of trends affecting the business, but are prevented from doing so by being placed into a long queue due to limited available resources. And while some low code/no

code tools are available for use with some of the data warehousing solutions, especially cloud-based products, this is not universal enough to include the full breadth of organizational data needed for a complete insightful analysis.

"…Enterprises need to examine all of these features in a unified solution to best take advantage of the potential benefits that enhanced analytics can provide in business operations, customer insights, and product directions. While the market continues to evolve, enterprises should look at the Cloudera suite of products as a way to achieve most of the capabilities necessary to run a modern data driven organization....."

Solving this challenge is a difficult undertaking. Cloudera has pioneered an ability to combine data using its Data Flow product which standardizes development and deployment of data flows. It simplifies and accelerates onboarding and managing of data sources and provides an ability to handle data in motion, ingesting content from a wide assortment of sources that can then be routed to any destination. Cloudera's Data Platform (CDP) creates a uniform base from which to power analytical solutions, including in a hybrid cloud environment. In addition, an ability to analyze data of virtually any type using Cloudera's Stream Processing enables faster time to value of the data as analysis takes place in near real time. By deploying tools that are low code (with Stream Builder), it opens up analysis to many more people who aren't data scientists or professional programmers. And once the data has been processed, it can then be incorporated into Cloudera's CDP data lakehouse for permanent storage.

We estimate that this solution can provide a significant improvement in speed of operations – often by as much as 50% or more. Further, with reduced technical barriers, it can not only speed development of analysis functions by as much as 75%-80% over traditional coding methods, but also opens up data analysis to a much wider audience of enterprise users.

This enables organizations to analyze data for immediate use supporting critical business events such as security incidents, operations anomalies, manufacturing problems, fraud detection, etc., where delays cause can significant disruptions and costs. Such a solution enables an organization to significantly enhance its digital transformation journey by embedding high quality real time insights into analysis of core business processes, while also enabling new insights that can power business initiatives much more quickly than with traditional analytical processes.

Bottom Line: With an increasing use of hybrid data locations and multi-cloud solutions, it is imperative that enterprises be able to run needed analytics on all of its dispersed data in order to stay competitive. Further, while many analytics solutions exist that are tied to specific data products (e.g., hyperscalers, dedicated SaaS data warehousing), the ability to ingest data from a wide variety of sources is lacking in most of these products. Finally, the need for near real time analysis by a wide variety of business users requires a solution to create data analysis functions using low code/no code capabilities. Enterprises need to examine all of these features in a unified solution to best take advantage of the potential benefits that enhanced analytics can provide in business operations, customer insights, and product directions. While the market continues to evolve, enterprises should look at the Cloudera suite of products as a way to achieve most of the capabilities necessary to run a modern data driven organization.

J.Gold Associates, LLC.
6 Valentine Road
Northborough, MA 01532 USA

Phone:
+1-508-393-5294

Web:
www.jgoldassociates.com

Email:
info@jgoldassociates.com

Research, Analysis, Strategy, Insight

J.Gold Associates provides advisory services, syndicated research, strategic consulting and in-context analysis to help its clients make important technology choices and to enable improved product deployment decisions and go to market strategies.