Cloudera Builds on Its CDP Data Platform, Embraces Private Cloud, Streaming

RFG Perspective: Cloudera Inc. is expanding its total available market (TAM) by building on its Cloudera Data Platform (CDP) – by entering the market for managed streaming and by making CDP available for private clouds, multiple public clouds and hybrid cloud deployments. The firm's next goal will be making CDP more accessible to non-technical business users and supporting streaming data for cloud-native environments.

Cloudera is planning to take its data management business in new directions – supporting public, private and hybrid clouds with an extensible data platform for the enterprise – and positioning itself to enter the emerging market for managing streaming data. As enterprises accelerate cloud migrations -- analyzing more of their data in the Cloud and at the Edge, they will need end-to-end data discovery, data security, and data governance to support the new business models.

In January 2019, Cloudera Inc. merged with Hortonworks Inc. – a company known for its Hadoop platform for data warehouses and data analytics. From the start, the merged Cloudera company focused on building a merged foundation for unified data management called the Cloudera Data Platform (CDP).

Bringing the two companies' data platforms together (including many open-source projects, including Hadoop, Hive, Impala, Kafka, Spark, and NiFi) into a single offering was a major accomplishment in 2019. Now, CDP itself is expanding its management of data in public clouds, private clouds, and hybrid clouds – and support end-to-end data security and governance policies.

It's a tough job in a tough year: The task at hand is for Cloudera – with an annualized recurring rate (ARR) of $739 million – to leverage CDP for many usage scenarios. That includes delivering CDP for use with multiple public clouds; for hybrid cloud deployments; and for private clouds, where data is secured and managed in an on-premises data center because of security, control or compliance factors. Cloudera's strategy was outlined at its recent Cloudera Now online conference. We believe that Cloudera is aiming to grow its top-line revenues to $1 billion, expanding its global business. CDP runs across the hybrid cloud, but that has the effect of making Cloudera attractive as a potential acquisition for a much larger company.

Importance of the SDX Engine

CDP provides a common data plane that moves corporate data from one resource to another – from the core to cloud to edge. At the center of CDP is the Shared Data

Experience (SDX) engine, which applies policies to secure and govern data, including on-prem and off-prem database instances.

To do that, SDX works with metadata – the data about the data – to enforce data security governance policies across hybrid clouds, from core to cloud to edge. However, we note here that customers cannot buy the SDX engine separately from the CDP offering that the SDX engine manages.

One more important piece of the strategy: the CDP platform is based on open-source code. This open-source foundation supports DevOps and DataOps for cloud-native environments – whether those workloads run on-prem or off-prem. Customers have the choice of applying open-source AI/ML, Kubernetes orchestration and storage software to cloud-native software developed for the hybrid cloud. Open-source software fits with preferences by many corporate DevOps groups to develop cloud-native software with containers, the Python and R programming languages and Kubernetes orchestration.

A Bit of History

The CDP platform has its roots in data warehouses and data-lakes that provide what's widely known as a "single version of the truth" by analyzing data that is housed inside an enterprise's data centers. Now, Cloudera's strategy is much larger in scope – encompassing both public and private cloud deployments.

By building out CDP as a unified data platform for customers, Cloudera can grow its data-centric franchise within enterprise accounts. Cloudera is well-positioned to take its product portfolio to entirely new sets of customers around the world, working with tech partners to extend its reach in the Americas, Europe/Middle East/Africa (EMEA) and Asia/Pacific, including Japan.

Into the Clouds – Public, Private and Hybrid

Data management in public clouds is a vital step in Cloudera's hybrid-cloud strategy – and that was shown when Cloudera delivered CDP for Public Cloud in late 2019 – providing support for cloud migration to cloud service providers (CSPs), including AWS, Azure and, soon, Google Cloud Platform.

Now, Cloudera is addressing the recent rise in customer's private-cloud deployments with its Cloudera Data Platform (CDP) for Private Cloud, first shipped in August 2020. This product is designed for enterprise customers who want the

cloud's flexibility and agility – while supporting the availability (RAS) and security associated with enterprise data centers.

The Importance of Private Cloud for Enterprises

Private cloud options are attracting customers' attention – as public cloud data gets repatriated to on-prem sites for reasons of security, availability and compliance with local or regional governmental regulations. Reverse cloud migration to on-prem locations can save costs for companies that found that large datasets running on the public cloud can be costly to maintain and operate.

In the overall marketplace, private cloud is rising in importance at many customer sites, which want to maintain granular control over data sources that feed advanced analytics and data sources that support mission-critical applications. CDP itself is based on open-source, to support DevOps as they build out cloud-native environments – whether they run on-prem or off-prem.

What's Next for CDP: A Focus on Streaming Data

Next, Cloudera is looking at CDP management of streaming, real-time data – which Cloudera views as generating the most data over the next five years. Streaming is a fast-growing source of cloud-native data worldwide and is expected to grow even more as enterprises try to analyze what is occurring with contractors and employees working from home (WFH) or other non-company sites.

Streaming data is already becoming a big contributor to enterprise data – although much of that is object data that remains uncategorized. That is because the streaming data, once it is stored and at-rest, may not yet be tagged with metadata that would support advanced data analytics.

However, transactional and streaming/stateless data are increasingly coexisting inside customer sites. One reason this is happening is the marriage of transactional data with consumer-oriented ecommerce in digitally transformed businesses. These two types of data – transactional and streaming data – can be connected via Apache Kafka, for example, with more traditional enterprise data center data-stores.

Data Democratization and Personas

The "people" side of analytics at customer sites is changing, too, in a process called data democratization – making data more widely available for purposes of analytics within business units. This is another reason why CDP is addressing both private clouds and public clouds – or their combination in hybrid clouds. There are IT users

of CDP – but there is a growing base of users based in customers' lines of business (LOB) organizations.

Data scientists are incredibly important when analyzing patterns in the data throughout the enterprise. But it turns out there is a range of "personas" involved in finding and using actionable data that improved business outcomes. Titles include data architects, data engineers, data scientists, business users and end-users. Importantly, many of these roles are not hired by IT, but rather by business managers within lines-of-business (LOBs).

Looking for Faster Growth

Cloudera is focused on catching the "wave" of digital change as enterprises update their company's data silos. The silos stemmed from corporate departments that each had their own IT technologies – or from mergers and acquisitions (M&A) that combined two companies without blending their data in a consistent way. So, Cloudera is tuning its go-to-market (GTM) plan to reach out to business managers, who drive much of the data-based spend in organizations.

Many of Cloudera's competitors will bring AI and automation to the marketplace. Product differentiation, ease of use and customer experience will be major factors in the adoption of data-related products. Cloudera has built a strong platform for data governance in the hybrid cloud era – including customers' multi-cloud deployments From here on, the success of Cloudera's strategy will come from successful execution of its technical plans – and from broadening its marketing messages to reach more business customers around the world.

RFG POV: Before their merger, Cloudera Inc. and Hortonworks Inc. were able to promote their products by explaining a largely technical story. Its end-to-end data management, security and governance technology will appeal to IT organizations that must implement their cloud-migration plans, redistributing data into the cloud and the edge. The merged Cloudera company has a broader charter than before, extending more deeply into customers' lines-of-business (LOBs) and corporate departments. Now, the story must be told in business terms. Cloudera's CDP is well-positioned to tap that larger opportunity but must make its story clear to business users.

Additional relevant research and consulting services are available. Interested readers should contact Client Services to arrange further discussion or interview Jean S. Bozman, President of Cloud Architects Advisors LLC, who authored this report.