REPORT REPRINT

Cloudera strives to be a thought leader in asia's data economy, lays foundation with CDP

OCTOBER 16 2019

By Agatha Poon

Built on the notion that hybrid is for now and in the future and that it enables analytics from the edge to ML/AI, regardless of IT environment, Cloudera has recently unveiled a new, cloud-native data platform, called Cloudera Data Platform, for public cloud.

Introduction

As all eyes turn to Asia as the next battleground for large-scale 5G deployments, with operators in Australia, Japan, South Korea and China gearing up to ignite innovation with cloud computing, IoT and ML/AI, Cloudera is taking notice. Positioned as an enterprise data cloud company, the vendor says the combined engineering team has worked on a new, cloud-native architecture upon the completion of technological and operational integration with Hortonworks. With hybrid/multi-cloud deployment, security and governance, and openness in mind, Cloudera starts a new chapter in the data world with its Cloudera Data Platform (CDP). The company has been working with selected customers for the preview version of CDP running on public cloud since August and that is now commercially available for business organizations. CDP for datacenters will be available this year and CDP for private cloud is planned for late in 2020.

451 TAKE

Cloudera has delivered on its promise of a data platform that maximizes the strengths of the combined entity. And its flexible deployment model, with CDP running on a public, private or on-premises environment, is particularly well suited in Asia-Pacific because individual economies are still at different stages of cloud maturity. Doubling down on open source technology positions Cloudera well for mainstream enterprise adoption when the time comes. Overall, we view this as a positive move for the company. It will be interesting to see if Cloudera can build on this start to penetrate further into Asia's mobile-first economy.

Context

Headquartered in Singapore, Cloudera has worked with customers at all levels for digital transformation projects since 2016. Hortonworks, on the other hand, shows strong interest in tapping big-data opportunities among enterprises, including government users, across industrial verticals and since 2017 has established a local presence. On the completion of the merger, Cloudera has increased its headcount across business functions – sales and marketing, product engineering, customer support, and training and advisory – in Asia-Pacific. Although the company declined to disclose aggregate regional headcount, it indicates that 40% of its employees are in product engineering, and pre-sales professionals have a good depth of technology knowledge. Additionally, it maintains a technical support center in India and is further investing in R&D to accelerate technology innovation with a commitment to 100% open source.

Given that both Cloudera and Hortonworks have similar clientele – going after both enterprises and service providers, Cloudera says it already captured cross-selling opportunities during the period of pre-integration with Hortonworks. The completion of the merger allows the new Cloudera to focus on deepening customer support with new capabilities and features and prioritizing the product roadmap to tune to the same wavelengths, so to speak, of their respective customers in Asia-Pacific. On the vertical front, the company is zeroing in on companies in the retail, manufacturing, financial services, telecommunications and public sectors because cloud migration has grown past the novelty phase and the appetite for innovative technologies such as IoT, blockchain and ML/AI is growing. Cloudera partners a great deal in Asia-Pacific – global SIs and technology providers in particular – to drive high-touch engagements. Some of its key partners include Accenture, AWS, Deloitte, Hitachi, Microsoft, NTT Data and Tata Communications.

With CDP, Cloudera is doubling down on open source. Having a combined team of more than 700 engineers contributing to the open source projects, all software products will be using open source codes with no vendor lock-in, Cloudera reiterates. There is no doubt of the company's technical expertise; the next strategic move is to make its software enterprise-grade. To that end, Cloudera claims to have witnessed significant growth, with IoT and connected car use cases coming in sight in the manufacturing and automotive sectors, and it is building a healthy pipeline among government customers. Investing further in capabilities such as adaptive scaling and unified management reflects enterprises' wishes to bring simplicity out of the complex hybrid IT environment.

Product

As the majority of customer data (nearly 70%) is still running on-premises, Cloudera says CDP is built on the notion that hybrid is for now and in the future, and it enables analytics from the edge to ML/AI, regardless of IT environment (public, private, on-premises and hybrid). In reality, the company agrees that CDP is more ML than AI-driven; by which it is primarily used to automate and customize workloads to create algorithms, and then automate decisions that used to be made by humans. This in turn presents huge opportunities for enterprise customers in terms of controlling operational costs.

CDP is a combined capability between Cloudera and Hortonworks, which also led to a lot of re-engineering to identify overlapping capabilities and optimize individual strengths. For example, data warehouse was a core capability for both parties, but each had it from a different angle and for different use cases. There are blended technologies to create a path for customers to extend their respective investments in Cloudera Distribution of Apache Hadoop (CDH) and Hortonworks Data Platform (HDP).

Initially delivered as a PaaS optimized for public cloud deployments, CDP for public cloud will now be running on AWS and Microsoft Azure, followed by Google Cloud Platform. Whether CDP will be extended to other public cloud providers will be driven by customer demand and cloud market share, Cloudera notes. Additionally, CDP for datacenter (on-premises bare-metal deployments) and CDP for private cloud will come online in due course.

At the core, CDP for public cloud provides two new services – data warehouse and machine learning services – for creating self-service applications to support different data users (business analysts and data scientists). There is a data hub service, which is used to create virtual private clusters for developers to build custom business applications with a unified open source distribution (Cloudera Runtime). The company says Data Hub and Cloudera Runtime is the first post-merger offering delivered as PaaS to help ease the tension between business and IT folks and provide a rich cloud-native experience.

Using a unified management console, customers are able to manage, monitor and orchestrate all the CDP services with a single-pane view across all environments. In order to have consistent data security, governance and control across hybrid and multi-cloud environments, the company enables a shared data experience with Cloudera SDX.

In Asia-Pacific, business organizations are slowly but clearly moving to a new generation of data analytics platforms that are highly distributed and open standards-based, Cloudera says. With CDP, adaptive scaling is native to the platform. Leveraging capabilities from both Cloudera and Hortonworks, CDP enables intelligent migration and ties into cloud bursting – along with security policies go with it.

Strategy

Aside from being fully committed to open source technology, Cloudera indicates that a new licensing policy went into effect on September 1 by which the company has made the binary distribution available with a license. Under the new licensing scheme, which is closely aligned with the proven Red Hat open source model, customers will need a subscription agreement for enterprise support, which enables customers to get access to the source code and all product binaries that have been tested and verified by Cloudera. Trademarks embedded within the binaries and Cloudera hosted source codes will remain protected, the company notes. Accordingly, the actual billing will take effect on October 24. The company says it will consolidate and transition a small number of projects currently licensed by Cloudera under closed source licenses to open source licenses. Consequently, all product sources will adhere to one of the two Open Source Initiative (OSI) approved licenses – either the Apache Software license (ASL) version 2 or the GNU Affero General Public License (AGPL) version 3.

Competition

Because Cloudera has extended its platform capability and focused on delivering higher-value use cases such as ML, AI and advanced analytics, the company believes that its new competition comes from the likes of Amazon SageMaker and other data science platform vendors. SAS, for example, is looking to exploit its expertise in machine learning and other data science realms with a cloud platform it calls Viya. TIBCO is seeking to serve enterprise data science and analysis needs on-premises and in the cloud. When it comes to delivering data science and machine learning as a service, Google TensorFlow is gaining attention. H20.ai is looking to make data science enterprise-ready using open source technology.

SWOT Analysis

STRENGTHS

Cloudera has made efforts to rationalize platform capabilities and accelerate integration. The making of CDP seems to be at the right time to broaden its market reach from ML-driven datacenters all the way to the edge.

WEAKNESSES

While the company has ramped up its data analytics capability with CDP, there is still room for improvement when it comes to demonstrating its prowess in AI with proven use cases.

OPPORTUNITIES

With business organizations seeking a more holistic way to analyze data coming from different sources, there are great opportunities for providers that can focus on delivering a unified experience in managing a variety of workloads in various security and performance sensitivity.

THREATS

Leading cloud service providers have invested considerably in boosting their ML/AI capabilities with integrated offerings and will continue to do so. Being part of the cloud vendor ecosystem, Cloudera may need to tread carefully when it comes to enabling CDP in the public cloud environments.