SOLUTION BRIEF

HORTONWORKS AND MICROSOFT: STRATEGIC PARTNERSHIPS IN THRIVING HADOOP ECOSYSTEMS

The challenges of big data are well worth the opportunities. The ability to gain insights from new forms of data (and previously difficult to work with data) is now a matter of choice in architected deployment options. With the democratization of big data, too, everyone can enjoy its benefits. The strategic Hadoop ecosystem partnership between Hortonworks and Microsoft delivers a new set of architected solutions with key benefits for more companies.

The strategic partnership between Hortonworks and Microsoft demonstrates a shared vision for democratizing big data since first announced in 2011. Because of the combined capabilities of both companies in data technologies, this partnership of more than two years has significant benefits for the enterprise.

DEEP ENGINEERING

As part of the relationship the engineering teams from both companies have worked together to achieve substantial engineering integration. First, both companies worked together to port Apache Hadoop, a Linux based technology, to run on Microsoft Windows Servers. Subsequently Hortonworks released the Hortonworks Data Platform (HDP) for Windows with Microsoft's full backing. During that time Hortonworks introduced the Stinger initiative with Microsoft as a key contributor. The Stinger initiative was a community effort to improve Hive performance by 100x as part of the Hadoop 2 YARN architecture. This was no minor feat: Microsoft engineers not only contributed expertise in high performance vectorized query processing from the advanced SQL Server engine to Hortonworks' engineers, but consequently to the open source community, too. Further collaboration has expanded the integration and compatibility with Microsoft Office and system monitoring tools that still continues today.

This shared vision means that both Hortonworks and Microsoft depend on one another to achieve their mission, making Hadoop readily and easily available to the Microsoft community.

Whether it's Hadoop on-premises, in Microsoft Azure, or both with a hybrid-cloud architecture, both Hortonworks and Microsoft will ensure that the world's most widely used productivity software -- MS Office -- can access all of the data available and deliver a rich user experience that's familiar and insightful. Equally as important is that both get a big step closer to enabling as many people and organizations as possible with an easily accessible and fully robust Hadoop ecosystem with HDP that can operate on Windows Servers, HDP in Azure as IaaS, HDInsight in Azure as PaaS -- all with MS Office interactivity.

BUSINESS DRIVERS

As companies evaluate longer-term Hadoop implementation strategies, we are seeing a noticeable increase of deep discussions pertaining to the adoption of a Data Lake Strategy, with the roles and benefits of Hadoop clusters architected on-premises, multi-premises, cloud-based, and hybrid-cloud. The overriding business driver remains reducing infrastructure management costs through optimized data platforms. However, this is to be expected following the wave of business justification in adopting Hadoop for business advantages and opportunistic value as optimization typically follows innovation.

The business drivers for Hadoop implementation strategies are balanced with technical, operational, and compliance feasibilities. Data may be ingested on-premises or in the cloud, based upon proximity to data sources or consumers due to the costs associated with data transfer. Hybrid-cloud architectures are being recognized as the optimal architecture for long-term Hadoop and Data Lake strategies, with efficiencies sought in existing resources, skillsets, and monitoring. Companies are recognizing these advantages in the Hortonworks and Microsoft Azure architected solutions.

© 2014 Radiant Advisors. All Rights Reserved.

SOLUTION BRIEF

HORTONWORKS AND MICROSOFT:

ARCHITECTED SOLUTIONS

Enabling Control, Flexibility and Customization

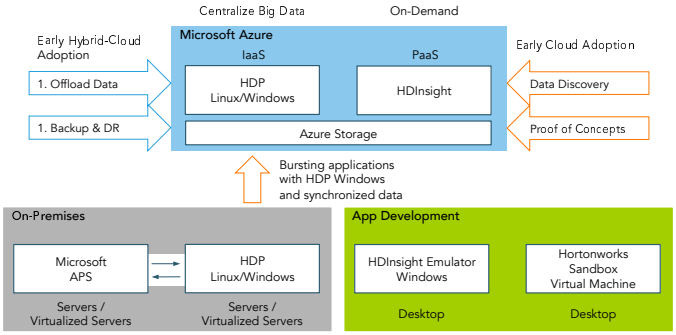
The HDP for Windows has enabled IT departments to leverage its existing proven hardware, operating system management, and process capabilities to deploy Hadoop clusters for customers on-premises. This includes the capability to run in private cloud infrastructure. Companies further operationalize the HDP cluster by backing up the on-premises HDP cluster to the Microsoft Azure cloud for further protection from localized disaster recovery and business continuity.

For public and hybrid-cloud deployments, Microsoft Azure offers a smooth transition to the cloud, or a transparent hybrid-cloud Hadoop architecture with on-premises HDP for Windows cluster management with Microsoft Systems Center 2012. Hadoop applications developed either on-premises or in the cloud can be assured of operating in both the on-premises HDP on Windows clusters and Azure HDInsight, and through hybrid-Hadoop architectures, if needed.

If there is a need to run HDP on Linux, Microsoft Azure will support that with its IaaS offering -- with full control to customize as needed. In the IaaS model, running a Hadoop cluster will incur charges for continuous time that it is up and running, regardless of usage activities. IaaS is rented infrastructure and data is stored in the Hadoop Distributed File System (HDFS) to be accessible.

Enabling Agility and Discovery

Companies find Microsoft Azure's HDInsight service an efficient architecture that is easy to use. In the efficient PaaS model, HDInsight as a managed service will provision a Hadoop cluster for use, and load data that is stored in the low cost Azure Storage. In this approach the HDFS acts as a data cache for the Hadoop cluster. Also, the number of nodes and amount of parallelism can be assigned at provisioning time, rather than being fixed as in an existing IaaS Hadoop cluster.



Microsoft Azure HDInsight is a managed service that enables business users to quickly perform data discovery in both self-service and on-demand manner. Leveraging data that is easily stored into Azure Storage, the business user can quickly provision and load a HDP cluster of any size and work with the data through MS Office or other discovery and visualization tools. Finally, HDInsight can be shut down just as easily when done, minimizing costs while the original data is still stored in Azure Storage.

For gaining familiarity and developing Hadoop applications, the Hortonworks Sandbox is a downloadable virtual machine of HDP for individual desktops. There is also the HDInsight Emulator that can be run on developers' desktops to ensure a smooth transition and deployment to Microsoft Azure HDInsight. This is considered to be lower development costs with existing desktops serving as a single node environment. Developers can also run a single node of HDP on Windows Server or use Windows 8 remote connection to a Windows Server running HDP.

Integrated for Big Data and Analytics Unification

A key for users is to have ubiquitous access to data that is both in the Hadoop cluster and relational analytics optimized databases -- like the Microsoft Analytics Platform System (APS) -- for a unified discovery and analytics experience. Therefore, Microsoft has developed Polybase that provides a federated data access experience to data both in the APS and HDP cluster. This is accomplished with the APS Data Movement Service (DMS) that can move data from HDP to APS transparently when requested.

© 2014 Radiant Advisors. All Rights Reserved.

SOLUTION BRIEF

HORTONWORKS AND MICROSOFT:

Most important will be all of the enhancements brought to MS Office that enables analytics, advanced and geo-spatial visualizations, and collaboration for users working with data. Initially, this began with an ODBC driver to Hive -- which is why Hive performance was so important. Microsoft has enabled Excel for newer analytics, visualizations, and Hadoop usage with Power Pivot, Power View, and Power Map. Self-service is enabled with Power BI through Power Query, and collaboration for users leverages both the new Power BI sites and SharePoint.

ADOPTION TRENDS AND LEADING USE CASES

While some companies consider it "low hanging fruit," we point out that a low risk opportunity to non mission-critical data is a good first step in gaining experience with hybrid-cloud Hadoop architectures. Offloading historical data or as-needed data sets from on-premises data warehouses allows companies to control data warehouse footprint costs and gain performance. This offload capability lowers storage and infrastructure costs for lower service-level agreement data access. Companies are initially copying data to a cloud-based Hadoop cluster or storage and gaining experience with data unification configuration, performance, and systems monitoring before purging data from the on-premises data warehouse. This safer initial approach also allows for gaining experience with structured data and schema within Hadoop and SQL access. Cloud-based redundancy or back-ups of on-premises Hadoop cluster are also being implemented as a beneficial, low risk opportunity to gain hybrid-cloud Hadoop experience.

While on-demand clusters in the cloud for discovery have gained acceptance, the necessary governance processes are only now being adopted. The use case and value proposition has been validated for users to create and utilize cloud-based Hadoop-as-a-Service, such as HDInsight, without requiring much technical support for agile discovery. Companies are focused on taking the results of discovery environments, adding them to existing Hadoop cluster on-premises, and "registering" it.

The use cases for bursting scenarios are being further analyzed with data synchronization between on-premises and cloud. In order for the Hadoop application to be transitioned to the cloud-based cluster, it requires the same data in both places with some degree of synchronization and latency to be resolved. Dual-feeding data to both locations or synchronizing does incur data transfer costs that are calculated along with bursting frequency, size, and nominal on-premises cluster sizes. While the technology is available today for HDP with Microsoft Azure IaaS and PaaS HDInsight to transparently manage bursting applications, companies are testing bursting as the next adoption wave.

KEY BENEFITS

Organizations experienced with operating the Microsoft environment are finding adopting Hadoop from Hortonworks much easier in leveraging existing resources, skills, and processes.

Microsoft Azure is a competitive cloud offering for the architecting Hortonworks Hadoop Data Platform with both the freedom of IaaS deployments and the efficiency of big data managed services with HDInsight.

Hybrid-cloud architecture is more easily achievable with the high compatibility between Hortonworks on-premises and cloud-based with Azure IaaS and PaaS through Systems Center and Ambari.

On-demand discovery and self-service are realized by business users and developers with Microsoft Azure HDInsight and Hortonworks Sandbox.

Microsoft Office enhancements have made Hadoop more accessible for everyone through Hortonworks HDP.

Radiant Advisors

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.

© 2014 Radiant Advisors. All Rights Reserved.