

LA  
INFRAESTRUCTURA  
DE **COMPUTACIÓN**  
NECESARIA PARA  
EL **ANÁLISIS**  
PREDICTIVO

Las organizaciones modernizadas aprovechan los beneficios transformativos del análisis predictivo para impulsar el avance de su negocio. El análisis predictivo le aporta conocimientos e información valiosa que pueden marcar la diferencia entre ganancias y pérdidas. Utilizan el modelado, el aprendizaje automático y la inteligencia artificial para analizar los datos actuales y realizar predicciones sobre el futuro. Cada sector puede beneficiarse del análisis predictivo. Quienes ignoran el potencial aquí presente están despilfarrando una gran oportunidad para crecer o crear mejores experiencias.

El notable progreso reciente en el análisis predictivo y el aprendizaje automático basados en la IA brinda a las organizaciones la capacidad de liberar un gran valor de los datos que antes eran difíciles de descifrar. A través de la potencia de la infraestructura habilitada para la IA, puede aprovechar la ola de análisis predictivo/aprendizaje automático y obtener ventajas para el negocio que no solo lo mantendrán en el juego, sino también a la vanguardia. Las funcionalidades automatizadas de los servidores y el servidor adecuado para el trabajo le permitirán ir donde sea necesario. Este e-Book se centra en la manera en que la automatización de servidores, junto con los servidores adecuados para la carga de trabajo, permite a las empresas realizar grandes avances en el mundo del aprendizaje automático y el análisis predictivo.



Un informe reciente de investigación de ESG, [Three Transformational Compute Technologies Verified to Accelerate AI and Business Value](#), segmentó las organizaciones de acuerdo con su nivel de avance en las funcionalidades de inteligencia artificial. Los tres criterios utilizados como análisis comparativos fueron la automatización de servidores, los aceleradores y la infraestructura convergente/hiperconvergente. Disponer de funcionalidades automatizadas para los servidores ayudó a impulsar las organizaciones a la categoría más avanzada (Etapa 3). Aunque solo aproximadamente la mitad (55 %) de las empresas en la Etapa 1 desarrollan, implementan y ajustan los modelos de IA en producción para el análisis predictivo, un impresionante 74 % de las empresas en la Etapa 3 están adoptando estos objetivos<sup>1</sup>.

¿Por qué es importante? Hay más casos de uso para el análisis predictivo y el aprendizaje automático de los que aparecen en este e-Book. Estos son solo algunos ejemplos:

- Creación de motores de recomendación para los clientes a fin de mantenerlos comprando sus productos de la manera que sea pertinente para ellos.
- Las fábricas y las cadenas de suministro pueden utilizar análisis para predecir cuándo se producirán fallas y así eludir el tiempo de inactividad.
- Predecir el precio óptimo para vender un producto en función de las tendencias de datos anteriores, como el tráfico peatonal y las proyecciones meteorológicas.

Estos escenarios ni siquiera alcanzan a entrever lo que usted puede hacer con el análisis predictivo. Mediante la automatización de sus servidores, libera tiempo de los empleados para que se centren en el avance de las iniciativas de IA en lugar de cuidar la infraestructura. Así, su personal puede trabajar en la consolidación de la información en lagos de datos, utilizará aplicaciones como Hadoop y Spark para resolver problemas, y deducirá predicciones útiles de esa gran cantidad de datos que ha recopilado durante años. Y todo esto se ejecutará en servidores con una potencia de procesamiento, almacenamiento y potencia de memoria excelentes. Cuando priorice el análisis predictivo y la infraestructura moderna, estará más cerca de ser una organización en la Etapa 3. Imagine si pudiera aprovechar estos beneficios.

## Organizaciones en la Etapa 3:<sup>2</sup>

tienen **2,6 veces** más probabilidades que las organizaciones en la Etapa 1 de liderar la competencia en inteligencia comercial y análisis.

tienen **el doble** de probabilidades de lograr un tiempo de respuesta que genera valor más corto.

mejore la velocidad de la toma de decisiones con la IA a casi **el doble** frente a las organizaciones en la Etapa 1.

mejore la precisión de la toma de decisiones con la IA en casi **2,6 veces** frente a las organizaciones en la Etapa 1.

observe una reducción promedio de **2,7 veces** en los costos mediante la automatización de las operaciones o los procesos de negocios en comparación con las organizaciones en la Etapa 1.

## Etapas de madurez de la IA de ESG

### Etapa 1

(**42 %** de las organizaciones en el estudio): bajos niveles de automatización, uso muy limitado de aceleradores o poca o nada de infraestructura convergente/basada en HCI para la IA.

### Etapa 2

(**33 %** de las organizaciones): niveles moderados de automatización, algún uso de aceleradores o algo de infraestructura convergente/basada en HCI para la IA.

### Etapa 3

(**24 %** de las organizaciones): niveles elevados de automatización, amplio uso de aceleradores o alto uso de infraestructura convergente/basada en HCI para la IA.

# Aplicaciones que lo ayudarán a ganar en el análisis predictivo: Spark en ejecución en Hadoop

Hadoop es una plataforma de código abierto que proporcionará la base para sus actividades de análisis predictivo y aprendizaje automático. Cuando se ejecuta el análisis predictivo, su primer reto (después de determinar qué problemas de negocios desea resolver o las oportunidades de mercado que desea aprovechar) es familiarizarse con la manera en que se reúnen todos los datos pertinentes. La recopilación de todos los datos pertinentes le permite agrupar los recursos en pools, de modo que las predicciones generadas sean lo más precisas posible. El mayor beneficio de Hadoop es que es un sistema de archivos altamente distribuido que le permite almacenar cualquier dato en formato nativo sin transformaciones, lo que lo hace ideal para los datos estructurados, pero también para los datos no estructurados.

Algunos ejemplos de casos de uso de Hadoop:



Los comerciantes minoristas brindan un mejor servicio a sus clientela mediante el uso de la TI para ayudar a analizar datos estructurados y no estructurados con el fin de ayudarlos a adaptar sus ofertas con mayor precisión.



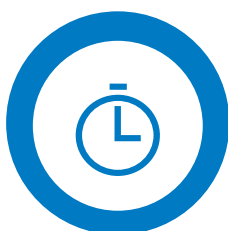
Hadoop proporciona datos a las aplicaciones que permiten a las empresas de servicios financieros impedir el fraude.



Las empresas de logística utilizan el análisis con tecnología de Hadoop para determinar si se requiere mantenimiento preventivo en sus equipos.

Es importante tener en cuenta que Hadoop se debe asociar con otra aplicación para ejecutar el análisis predictivo y el aprendizaje automático. A menudo, esa aplicación es Spark. Spark es un proyecto de código abierto con funcionalidades mejoradas. En Hadoop, tiene funciones de mapeo y reducción. La función de mapeo toma un conjunto de datos y lo convierte en otro conjunto de datos. Este nuevo conjunto desglosa los elementos en pares de clave/valor. La función de reducción luego combina estos pares de clave/valor en un conjunto más pequeño de pares de clave/valor. Adicionalmente, en Spark puede unir, agrupar en clústeres, filtrar y agregar o eliminar datos. Esto le ofrece oportunidades adicionales para manipular sus datos. Spark ofrece análisis de alta velocidad y tiene una biblioteca de aprendizaje automático incorporada. El motor de procesamiento en memoria de Spark garantiza resultados predictivos en tiempo real. Es importante recordar que Spark y Hadoop trabajan en asociación para ofrecer estos beneficios a su organización.

Funciones importantes de la implementación de Spark:



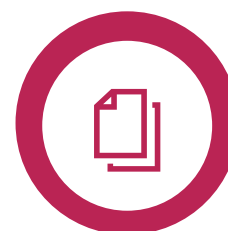
Consulta de datos en tiempo real



Una amplia comunidad de soporte



Rápido procesamiento de flujos de datos de baja latencia; ejecuta cargas de trabajo 100 veces más rápido que Hadoop<sup>3</sup>



Bibliotecas enriquecidas

# Se necesitan servidores completos y automatizados para ejecutar el análisis predictivo

Para aprovechar al máximo el emparejamiento de Hadoop y Spark, necesitará lo siguiente:

1. Servidores automatizados que liberen tiempo y recursos para fines de aprendizaje automático y que brinden información valiosa más rápidamente a sus clientes.
2. Procesadores con núcleos abundantes y potentes para analizar todos esos datos y ejecutar esos algoritmos.
3. Almacenamiento sustancial para asegurarse de que Hadoop se ejecute sin limitaciones.
4. Gran cantidad de memoria de servidor para permitir que el procesamiento en memoria de Spark funcione de manera óptima.

## Automatice sus servidores

¿Qué tiene que ver la automatización de los servidores con la ejecución de análisis predictivo y aprendizaje automático? Mucho, en realidad. Si reemplaza los servidores obsoletos por servidores modernos con funcionalidades de automatización, puede esperar lo siguiente:<sup>4</sup>

**43 %** | mayor rapidez en las implementaciones

**43 %** | mayor confiabilidad de los sistemas

**38 %** | mayor rapidez en las actualizaciones de las aplicaciones

**37 %** | menor tiempo dedicado a la administración manual y rutinaria de la infraestructura de TI

Esto se traduce en una implementación más rápida de información valiosa y más tiempo para que el personal se centre en entregarla. Usted desea que sus servidores de análisis se mantengan activos y se ejecuten increíblemente rápido, y que vuelvan a estar en línea cuando dejen de funcionar. Desea poder implementar los clústeres de Hadoop con eficiencia y actualizar Hadoop y Spark lo más pronto posible cuando sea necesario. Cuanto más automatice, menos tiempo gastará en procesos manuales de mantenimiento que puedan desacelerar su viaje hacia la obtención de valor a partir del análisis.

## Busque el servidor completo adecuado para Hadoop y Spark

Lo que necesitará para el trabajo de análisis predictivo con Hadoop y Spark es un servidor que lo tenga todo: procesamiento, almacenamiento y memoria con buenos resultados.

Los procesadores son el cerebro detrás de la operación, el análisis y la interpretación de todas las instrucciones que otras soluciones de hardware y software les envían. En el aprendizaje automático, el procesador está a cargo de ejecutar la lógica en el algoritmo. Esta es una gran responsabilidad; el conteo de núcleos es absolutamente importante. También es importante asegurarse de elegir la infraestructura que pueda admitir estas cargas de trabajo de procesamiento intensivo.

Spark recomienda el aprovisionamiento de al menos 8-16 núcleos por máquina a fin de analizar todos los datos necesarios para el análisis predictivo<sup>5</sup>.

Hadoop es una aplicación que demanda almacenamiento y que realiza el procesamiento en lotes a gran escala. Procesa grandes cantidades de datos en el rango de petabytes. Para ejecutar Hadoop, asegúrese de aprovisionar servidores que cuenten con almacenamiento suficiente. A pesar de que Hadoop sigue un modelo distribuido, cada servidor o nodo individual necesita un potencial de almacenamiento adecuado para los discos locales a fin de ejecutarse sin problemas. Es importante comprender que hay varios tipos de nodos en Hadoop. Este e-Book se centrará en las ofertas para el nodo en el edge, que actúa como el gatekeeper del clúster, y los nodos trabajadores, que impulsan el procesamiento real de los datos.

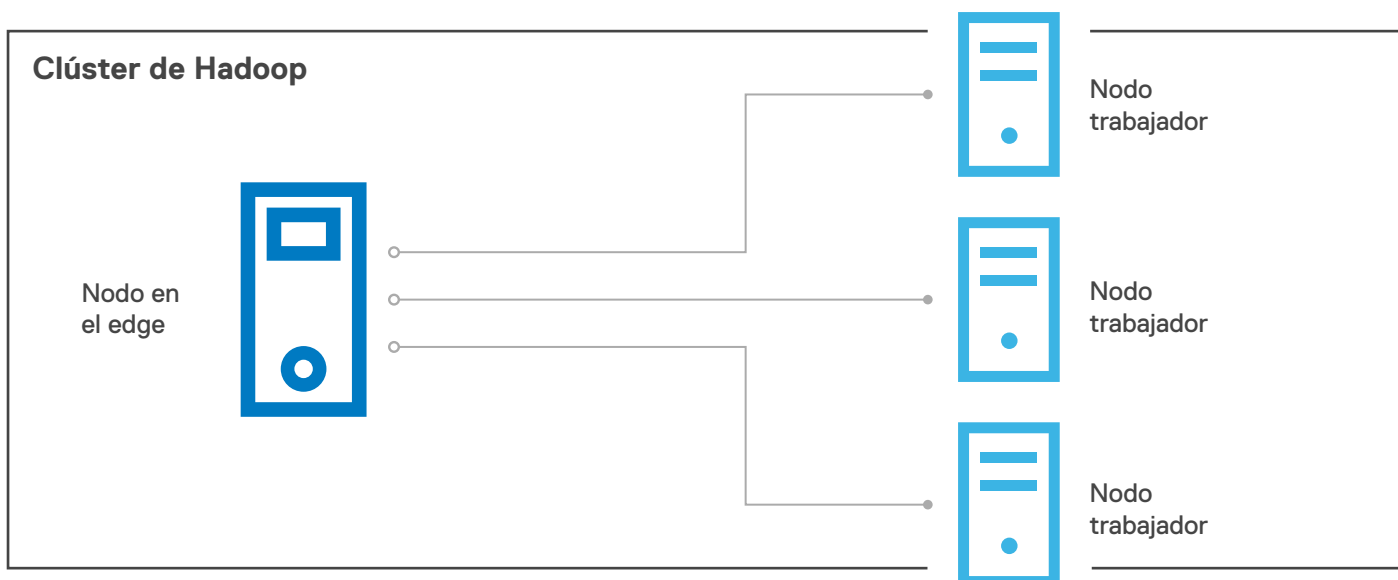
Por otro lado, Spark requiere mucha memoria. A pesar de que la recomendación mínima de Spark es de 8 GB de memoria por máquina, lo ideal es una cantidad mucho mayor. Algunos expertos recomiendan al menos 128 o 256 GB. Spark recomienda asignar el 75 % de la memoria disponible como máximo y dejar el resto para el sistema operativo y la caché del buffer<sup>6</sup>. La memoria es especialmente importante para esta aplicación, ya que almacena en caché los datos en la memoria en varias operaciones en paralelo.

En resumen, necesita servidores confiables y automatizados que puedan hacerlo todo cuando se trata de explorar el análisis predictivo y el aprendizaje automático. Siga leyendo para ver los servidores de Dell EMC diseñados para estas cargas de trabajo con uso intensivo de recursos.

# Servidores PowerEdge para sus necesidades de análisis predictivo

En primer lugar, todas las recomendaciones de productos que aparecen a continuación están precedidas por el hecho de que ofrecen opciones de automatización extremadamente atractivas. OpenManage, la plataforma de administración de servidores de primer nivel de Dell EMC, ofrece API RESTful. Las API le permiten generar scripts de muchos aspectos de la implementación, el mantenimiento y el aprovisionamiento de sus servidores. Puede escribir un script, configurarlo y olvidarlo. El portafolio de OpenManage también es compatible con la infraestructura de Ansible, que ofrece módulos que permiten la automatización de la TI en un escenario de múltiples proveedores. Estos módulos se ofrecen como scripts escritos previamente para una automatización optimizada<sup>7</sup>.

Para crear un clúster de Hadoop, necesitará varios tipos de nodos. El siguiente diagrama se centrará en un solo nodo en el edge y en varios nodos trabajadores. El nodo en el edge administra el acceso al clúster de Hadoop y actúa como el punto de entrada entre el clúster de Hadoop y la red externa. El nodo en el edge no está a cargo del trabajo de computación intensivo. Esa tarea recae en los nodos trabajadores. Los nodos trabajadores asumen la ardua tarea de procesar los datos y requieren una memoria y un almacenamiento generosos, así como poder de procesamiento. Para obtener información detallada acerca de Hadoop, puede leer más [aquí](#).



## Nodo en el edge

### PowerEdge R640

A pesar de que el nodo en el edge no está destinado al procesamiento intensivo de datos, aún es necesario contar con procesadores, almacenamiento y memoria potentes para mantener el clúster funcionando sin problemas. Dell EMC PowerEdge R640 es simplemente adecuado para esta tarea, ni más ni menos para esta función.

El R640 le permite escalar los recursos de computación con hasta dos procesadores escalables Intel® Xeon® de 2.ª generación con hasta 56 núcleos. Incluye hasta 24 DIMM, 12 de las cuales pueden ser NVDIMM o DCPMM, y hasta 7,68 TB de memoria<sup>8</sup>. Con PowerEdge R640, puede crear un pool de caché de NVMe y utilizar unidades de 2,5" o 3,5" para el almacenamiento de datos. Cree una configuración óptima de nodos en el edge con procesadores, memoria y almacenamiento escalables.



## Reconocimiento

**IT Brand Pulse: líder del mercado 2019 en servidores de montaje en rack:** "Los profesionales de TI volvieron a votar por Dell EMC como el líder del mercado de servidores de montaje en rack por tercera vez en siete años".



# Servidores PowerEdge para sus necesidades de análisis predictivo

## Nodos trabajadores

### PowerEdge R740

PowerEdge R740 es una plataforma adecuada para el análisis, con una memoria altamente expandible (hasta 7,68 TB) que hace que la ejecución de Spark sea muy sencilla. Escale para satisfacer las exigencias de capacidad con hasta 16 unidades de 2,5 in u 8 unidades de 3,5 in y hasta 128 TB de almacenamiento para mantener los clústeres de Hadoop funcionando de manera óptima. El R740 cuenta con la familia de procesadores escalables Intel® Xeon® de 2.ª generación con hasta 56 núcleos. Eso es lo que necesita para procesar grandes cantidades de datos.



### Reconocimiento

IDEA 2018 Bronze Award

### PowerEdge R740xd

Además de las funcionalidades del PowerEdge R740, el servidor Dell EMC PowerEdge R740xd agrega opciones de capacidad de almacenamiento extraordinarias, por lo que es ideal para las aplicaciones con uso intensivo de datos que requieren un almacenamiento significativo, como Hadoop. Seleccione hasta 24 unidades NVMe o un total de 32 unidades de 2,5 in o 18 unidades de 3,5 in y hasta 288 TB de almacenamiento. El R740xd con procesadores escalables Intel® Xeon® de 2.ª generación generó un 900 % más de conjuntos de consultas en pruebas recientes que el servidor de la generación anterior R720xd, y proporcionó los datos un 99,8 % más rápido. En comparación con nuestra versión más reciente de este servidor, el R740xd con procesadores escalables Intel® Xeon® de 2.ª generación alojó un 400 % más conjuntos de consultas y proporcionó los datos un 27 % más rápido<sup>9</sup>.



### Revisión

ServeTheHome: [análisis del servidor Dell EMC PowerEdge R740xd](#)



Opciones del nodo trabajador, comparación



Opción 1: R740



Opción 2: R740xd



Procesamiento

Hasta 56 núcleos

Hasta 56 núcleos



Almacenamiento

128 TB

288 TB



Memoria

7,68 TB

7,68 TB

## Arquitecturas preparadas para Hadoop

La pericia y la infraestructura marcan una diferencia al desarrollar un entorno de Hadoop<sup>10</sup>. Usted tiene la opción de reducir el esfuerzo de su parte y adquirir un plan validado acerca de cómo desarrollar los clústeres. Ready Architectures for Hadoop está diseñado para satisfacer los requisitos de análisis de datos, reducir los costos de desarrollo y mejorar el rendimiento. A pesar de que hay muchos casos de organizaciones que no cumplen con los plazos y tienen dificultades para tomar decisiones en cuanto a la configuración de sus plataformas de Hadoop, esa no tiene por qué ser la historia de su empresa<sup>11</sup>.

Dell EMC Ready Architectures for Hadoop son soluciones creadas para satisfacer todas las necesidades de análisis de datos. Dell EMC comenzó a desarrollar soluciones personalizadas de Hadoop en 2009 y cuenta con la pericia, las herramientas y las soluciones necesarias para impulsar una correcta implementación de Hadoop. El diseño de Hadoop de Cloudera ofrece los elementos clave de Hadoop dentro de una solución basada en el software Cloudera Enterprise y el hardware de Dell EMC, con opciones de servicio para su conveniencia. Los beneficios clave incluyen lo siguiente:<sup>12</sup>

- Aprovechar una solución ya optimizada y validada
- Superar la falta de pericia interna en Hadoop
- Acortar los cronogramas del proyecto y reducir los problemas de producción

Para obtener más información sobre Dell Ready Architectures for Hadoop, visite [nuestro sitio](#) o [contáctenos](#) hoy mismo.

## Servicios de migración de Hadoop

¿Está listo para aventurarse y migrar a Hadoop, pero no está seguro de cómo hacerlo? Los servicios de consultoría de Dell Technologies lo pueden ayudar. Si tiene dificultades para administrar las oportunidades creadas por el análisis de datos y las cargas de trabajo intensivas asociadas con grandes cantidades de datos, sepa que no está solo. Los servicios de consultoría de Dell Technologies ofrecen todo, desde una migración solamente de datos hasta una migración de plataforma completa, según las necesidades de su organización. Todo comienza con una evaluación del estado actual, lo que conduce al diseño de una solución de estado futuro. A continuación, realizamos una prueba piloto de la migración para probar y validar un entorno de muestra. Después de una prueba piloto correcta, la solución se implementa por completo<sup>13</sup>.

Para comenzar, [obtenga más información](#) sobre los servicios de migración de Hadoop o [póngase en contacto](#) con un experto de Dell Technologies Services.



# Historia de cliente: Epsilon

Echemos un vistazo al análisis predictivo en el trabajo. Epsilon es un líder de marketing detrás de los programas de fidelidad, correo electrónico y otros para las marcas de primer nivel. Utiliza una gran cantidad de puntos de datos a fin de personalizar los mensajes para sus diversos destinatarios, lo que garantiza los máximos resultados para los clientes. Como lo señala el director de TI de Epsilon, Robert Walden, “Todo lo que hacemos se centra en los datos y en nuestra capacidad de enviar el mensaje correcto a la persona adecuada en el momento oportuno”.

La velocidad es primordial para que la empresa continúe su trayectoria actual de duplicación del crecimiento año tras año. Esto requiere una infraestructura y un personal que pueda seguir el ritmo. Epsilon automatiza todos sus procesos de incorporación en la implementación de los servidores Dell EMC PowerEdge R740xd. El R740xd proporciona la flexibilidad, la escalabilidad y el rendimiento para cumplir con los requisitos de este negocio exigente. La ejecución de cargas de trabajo como Hadoop para aplicar IA y aprendizaje automático a sus procesos de personalización de correos electrónicos garantiza que Epsilon pueda cumplir con sus promesas.

El director de TI de Epsilon, Robert Walden, lo dice todo: “Estamos centrados en el éxito de nuestros clientes. Todo lo que hacemos en nuestra infraestructura y en nuestros entornos de aplicaciones es para facilitar ese éxito. Cuanto más eficaces sean las aplicaciones, las soluciones, los servicios, la infraestructura y el hardware, mejor será la posibilidad de alcanzarlo”.

[Lea](#) el caso de estudio de Epsilon.



[Vea](#) el video del caso de estudio de Epsilon.



Epsilon en el trabajo



Robert Walden, director de TI de Epsilon

# Conclusión

Con los servidores modernos y automatizados que cuentan con procesadores de alta potencia y generosas opciones de memoria y almacenamiento, su organización estará posicionada para aprovechar al máximo las múltiples oportunidades generadas por el análisis predictivo. Al mismo tiempo, dominar las ventajas y desventajas del análisis predictivo es clave para mantener la relevancia de la TI en su organización. Asíciase con Dell EMC para obtener soluciones que simplemente tienen sentido y eliminan las conjeturas de la introducción al aprendizaje automático y el análisis predictivo.

Para obtener más información, [póngase en contacto](#) con un representante de ventas de Dell EMC o visite [DellEMC.com/es-mx/Servers](http://DellEMC.com/es-mx/Servers).

<sup>1</sup> <https://www.dell EMC.com/es-mx/collaterals/unauth/analyst-reports/products/servers/esg-three-transformational-compute-technologies-verified-to-accelerate-ai-and-business-value-en.pdf>

<sup>2</sup> *ibid.*

<sup>3</sup> <https://spark.apache.org/>

<sup>4</sup> <https://www.dell EMC.com/resources/es-mx/asset/analyst-reports/products/storage/forrester-delivering-outcomes-by-automating-compute-infrastructure.pdf>

<sup>5</sup> <https://spark.apache.org/docs/latest/hardware-provisioning.html>

<sup>6</sup> *ibid.*

<sup>7</sup> <https://blog.dell EMC.com/es-mx/ride-ansible-revolution-dell-emc-openmanage/>

<sup>8</sup> <https://www.dell.com/en-us/work/shop/povw/poweredge-r640>

<sup>9</sup> <https://www.principledtechnologies.com/Dell/PowerEdge-R740xd-analytics-comparison-0719.pdf>

<sup>10</sup> <https://www.dell EMC.com/es-mx/solutions/data-analytics/hadoop/index.htm>

<sup>11</sup> <https://blog.dell EMC.com/es-mx/dell-emc-can-help-dive-successful-hadoop-projects/>

<sup>12</sup> [https://www.dell EMC.com/es-mx/collaterals/auth/white-papers/products/ready-solutions/Ready\\_Bundles\\_for\\_Hadoop\\_-\\_Solution\\_Overview\\_China.PDF](https://www.dell EMC.com/es-mx/collaterals/auth/white-papers/products/ready-solutions/Ready_Bundles_for_Hadoop_-_Solution_Overview_China.PDF)

<sup>13</sup> [https://www.dell EMC.com/es-mx/collaterals/unauth/offering-overview-documents/services/H16645\\_Big\\_Data\\_Migration\\_for\\_Hadoop\\_svo.pdf](https://www.dell EMC.com/es-mx/collaterals/unauth/offering-overview-documents/services/H16645_Big_Data_Migration_for_Hadoop_svo.pdf)