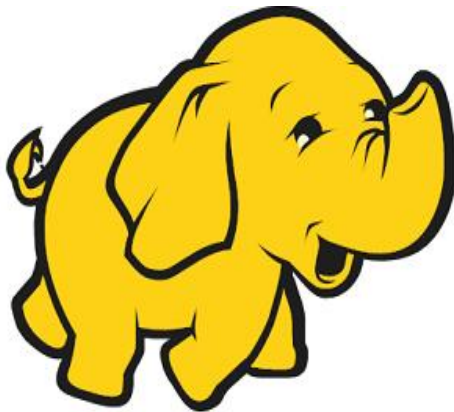


Servicio en la nube Azure HDInsight



Microsoft HDInsight



¿QUÉ ES AZURE HDINSIGHT?

Azure HDInsight es servicio de Microsoft Azure que nos provee una plataforma de **big data** como un servicio gestionado que le permite ejecutar las tecnologías **Apache Hadoop, Spark, HBase y Storm** sin gestionar la implementación ni la configuración. Puede implementar estas tecnologías en clúster de Windows o Linux y puede activar el clúster en cuestión de minutos. Puede procesar datos semi-estructurados y no estructurados y desarrollarse en Java, .Net, Python y más. También puede visualizar sus datos en Excel, Tableau o PowerBI.

¿CÓMO PROBAR EL SERVICIO AZURE HDINSIGHT?

Para probar el servicio **HDInsight**, se debe crear una nueva de Azure donde puede obtener créditos gratuitos que puede utilizar para crear un clúster de **HDInsight**.

IMPLEMENTACIONES EN AZURE HDINSIGHT

Incluye implementaciones de Apache Spark, **HBase**, Kafka, Storm, Pig, Hive, Interactive Hive, Sqoop, Oozie, Ambari, etc.

TIPOS DE CLÚSTER EN HDINSIGHT:

- **Apache Hadoop:** Ofrece almacenamiento de datos confiable con HDFS(sistema de archivos) con programación de MapReduce que procesar y analizar los datos en paralelo.
- **Apache Spark:** Framework de procesamiento en paralelo que admite el procesamiento en memoria. Mejora el rendimiento de las aplicaciones de análisis de Big Data, trabajos de Spark para SQL.
- **Apache HBase:** base de datos NoSQL en Hadoop, proporciona acceso aleatorio, la gran coherencia que maneja para Big Data no estructurados y semiestructurados; consultas de miles de millones de filas multiplicadas por millones de columnas.
- **Microsoft R Server:** Ofrece un servicio de servidor empresarial para hospedar y administrar procesos de R distribuidos en paralelo. Proporciona a los usuarios de Big Data, estadísticos y programadores de una gran cantidad de acceso a petición a métodos escalables y distribuidos para realizar análisis en HDInsight.
- **Apache Storm:** Aplicación distribuida de cálculo, se realiza en tiempo real para el procesamiento rápido de grandes volúmenes de datos. Es un clúster administrado en HDInsight.

¿ QUE LENGUAJES DE PROGRAMACIÓN SE UTILIZAN EN HDINSIGHT

Los clústeres de **HDInsight** clusters admiten varios lenguajes de programación, pero algunos de ellos no están instalados de manera predeterminada, De forma predeterminada, los clústeres de **HDInsight** admiten lo siguiente: **Java y Python.**

VENTAJAS DE HADOOP EN HDINSIGHT EN LA NUBE DE AZURE

- El aprovisionamiento automático de clústeres de Hadoop.
- Componentes de Hadoop de última generación.

- Alta disponibilidad y confiabilidad de los clústeres.
- Almacenamiento de datos eficaz y económico con Azure Blob Storage o Azure Data Lake Store.
- Integración con otros servicios de Azure, como Web Apps y SQL Database.
- Escalado de clústeres, poder crear más clusters cuando sea necesario
- Compatibilidad con redes virtuales. Los clústeres de HDInsight se pueden usar con Red virtual de Azure.
- Coste inicial bajo. Puede iniciar una prueba gratuita creando una cuenta en Azure HDInsight.

Título: Recuperación y acceso a la información

Autor: Marcos Alejandro Pariona Pariona

Publicado en Abril, 2017