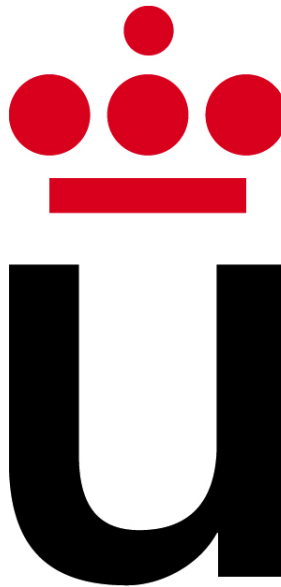


FORO DE PREGUNTAS

Análisis de Big Data

PREGUNTAS TEMA 3

José Ignacio Escribano



MÓSTOLES, 25 DE MARZO DE 2016

Índice

1. Preguntas	1
1.1. Microsoft se ha convertido recientemente en otro de los principales proveedores de servicios cloud públicos. Busque información sobre dichos servicios e indique cuál es la plataforma principal que ofrece Microsoft Azure para análisis de Big Data, resumiendo brevemente sus principales funciones.	1
1.2. ¿Qué tipo de servicios de virtualización y cloud computing podemos implementar con OpenStack, públicos o privados? ¿Ofrece OpenStack algún tipo de soporte para asignación dinámica de recursos computacionales dentro del cloud?	1

1. Preguntas

1.1. **Microsoft se ha convertido recientemente en otro de los principales proveedores de servicios cloud públicos. Busque información sobre dichos servicios e indique cuál es la plataforma principal que ofrece Microsoft Azure para análisis de Big Data, resumiendo brevemente sus principales funciones.**

Microsoft Azure es la plataforma que ofrece servicios cloud de Microsoft.

Microsoft Azure dispone de más de 50 servicios en la nube, entre los que destacan:

- Máquinas virtuales: aprovisionamiento de máquinas virtuales de Windows y Linux.
- Servicio de Aplicación: creación de aplicaciones web y móviles para cualquier plataforma y dispositivo.
- Bases de datos SQL: bases de datos SQL como servicio.
- Almacenamiento: almacenamiento en la nube duradero, de alta disponibilidad y escalable a gran escala.
- DocumentDB: bases de datos NoSQL de documentos como servicio.
- BackUp: copias de seguridad de los servidores en la nube.
- Aprendizaje Automático: potentes análisis predictivos en la nube.

La lista completa de estos servicios se puede encontrar en [1].

Con respecto a los servicios de Big Data de Microsoft Azure, tenemos HDInsight y Lote.

HDInsight permite el aprovisionamiento de clústeres Hadoop. Algunas de las características de este servicio son:

- Manejo de Apache Hadoop, Spark, HBase y Storm de forma sencilla.
- Escalabilidad hasta petabytes de datos en demanda.
- Procesamiento de datos estructurados, no estructurados y semiestructurados.
- Posibilidad de desplegar en Windows o Linux.
- Visualización de datos en Excel o en herramientas de Business Intelligence.

- Multitud de extensiones para distintos lenguajes de programación, entre los que se encuentran C#, Java o .NET.
- Integración con clústers de Hadoop en infraestructura propia.
- Personalización de clústers para correr otros proyectos de Hadoop, R, Giraph o Solr, entre otros.
- Procesamiento en tiempo real.
- Uso de Spark para análisis interactivo.

Lote es un servicio de computación por lotes (batch). Algunas de sus características son:

- Escalabilidad horizontal
- Programación de trabajos
- Copias intermedias de los datos

Respecto a los precios de estos servicios, sólo se paga por lo que se usa y Azure factura por minutos de uso. Una estimación del coste de los servicios se puede ver en su calculadora, que se puede consultar en [2].

Fuentes:

[1]: <https://azure.microsoft.com/es-es/services>
 [2]: <https://azure.microsoft.com/es-es/pricing/calculator/>
 [3]: <https://azure.microsoft.com/es-es/services/hdinsight/>
 [4]: <https://azure.microsoft.com/es-es/services/batch/>

1.2. ¿Qué tipo de servicios de virtualización y cloud computing podemos implementar con OpenStack, públicos o privados? ¿Ofrece OpenStack algún tipo de soporte para asignación dinámica de recursos computacionales dentro del cloud?

OpenStack es un proyecto para proporcionar Infraestructura como Servicio (IaaS).

OpenStack es un grupo de herramientas y APIs que permiten agilizar procesos a los administradores de un datacenter, en tareas como creación de máquinas virtuales, migración de nodos y manejo de recursos de la nube a personas poco experimentadas.

OpenStack permite la creación tanto de nubes públicas como privadas.

OpenStack consta de tres componentes principales: Compute, Networking y Storage.

El componente Compute, también llamado Nova, es el encargado de gestionar el ciclo de vida de las instancias de cómputo en un entorno OpenStack. Algunas de sus tareas incluyen la producción, el desmantelamiento y la planificación de máquinas en demanda.

Nova puede trabajar con tecnologías de virtualización. Para la tecnología del hipervisor las tecnologías disponibles son KVM y Xen, junto con Hyper-V, vSphere de VMware y la tecnología de contenedores Linux como LXC.

El componente Networking, también llamado Neutron, es el encargado de permitir la conectividad como servicio para otros componentes de OpenStack, como Compute. Provee una API para definir redes. Tiene una arquitectura que soporta muchas tecnologías populares de distintos fabricantes.

El componente Storage, también conocido como Swift, es el encargado de almacenar y recuperar datos vía API REST o cualquier API basada en HTTP. Es altamente tolerante a fallos, gracias a la replicación de los datos y a su arquitectura escalable.

Fuentes:

[1]: <https://www.openstack.org>

[2]: <http://www.openstack.org/software/releases/liberty/components/nova>

[3]: <http://www.openstack.org/software/releases/liberty/components/neut>

[4]: <http://www.openstack.org/software/releases/liberty/components/swif>

[5]: <https://es.wikipedia.org/wiki/OpenStack>

[6]: <http://hispavirt.com/2013/09/10/openstack-que-es-que-no-es/>

[7]: <http://hipertextual.com/archivo/2014/07/openstack-nube-virtualizac>