¿Qué nube debería utilizar en mi proyecto de Big Data? 5/52

Curso de Big Data en AWS

→|ARQUITECTURAS ...

Actualmente el mercado de Cloud Computing tiene varios actores compitiendo entre sí por atraer la mayor cantidad de clientes a sus nubes, encontramos Múltiples opciones como: Amazon Web Services, Azure, Alibaba Cloud, Google Cloud Platform, Oracle Cloud, Rackspace, Digital Ocean y Softlayer entre muchas otras.

Dentro de esta variedad de proveedores muchas veces es complejo tomar decisiones de cuál utilizar, el criterio para esta decisión puede estar dado por diferentes factores como:

1. **Costo:** Valor de los servicios que serán utilizados en el proyecto.
2. **Tipo de pricing**: Por demanda (por hora, minuto o segundo), subasta, reservado.
3. **Servicios:** Variedad de servicios provistos por el cloud provider. ¿Cuál servicio se ajusta mejor a mis necesidades?
4. **Ubicación:** Distribución de las regiones/zonas donde el cloud provider preste servicios por temas de latencia y experiencia usuario esto puede ser decisivo.
5. **Niveles de Servicio:** Consultar la documentación por servicio y los niveles ofrecidos de disponibilidad.
6. **Soporte:** Tipos de soporte, costo, tiempos de respuesta y nivel de soporte (basic, business, enterprise).
7. **Estudios de mercado:** Revisar los diferentes estudios de mercado, por ejemplo: el cuadrante mágico de Gartner, en los cuales se evalúan en diferentes aspectos los servicios provistos.
8. **Documentación:** Consultar la documentación de los cloud provider, muchas veces no es muy clara o está incompleta referente a sus servicios.

AMAZON WEB SERVICES & GOOGLE CLOUD PLATFORM

En los proyectos en los cuales se manejan grandes cantidades de datos y analítica, se utilizan **cloud providers** que tienen varias cosas en común:

DATALAKE/DATAWAREHOUSE

Son **almacenes electrónicos** que mantienen una gran cantidad de datos en bruto.

Redshift | Athena | Big Query

ALMACENAMIENTO POR OBJETOS

Los datos se manipulan como **unidades discretas,** llamadas objetos.

S3 | Cloud Storage

ETL: EXTRACT, TRANSFORM AND LOAD

Se referiere al **movimiento y transformación** de datos.

Glue | DataFlow

PROCESAMIENTO DE DATOS-HADOOP

Permite gestionar el estructura distribuida de los datos y su **procesamiento.**

EMR | DataProc

STREAM DE DATOS

Facilita la recopilación, el procesamiento y el análisis de datos **en tiempo real.**

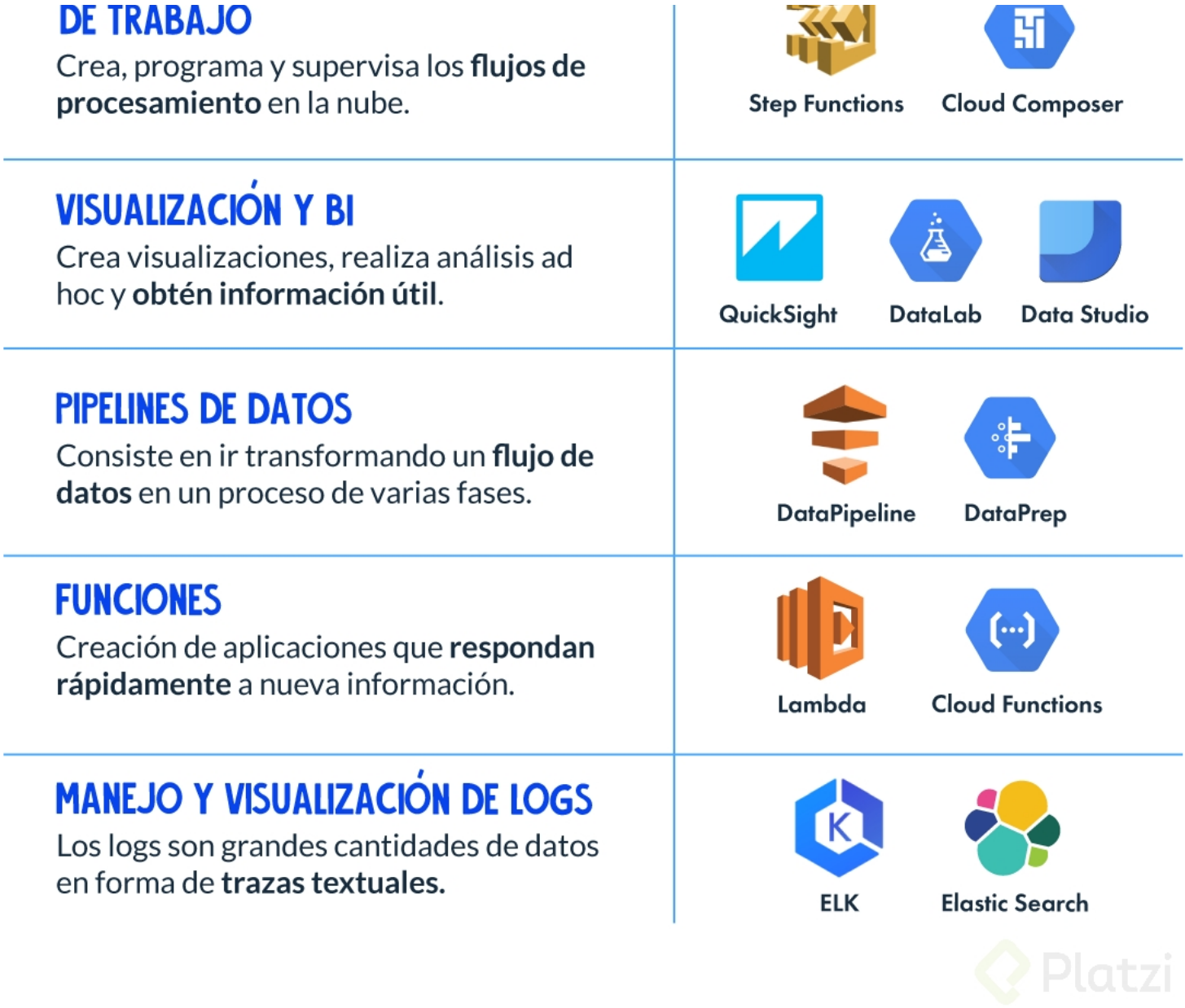
Kinesis | Pub/Sub

INFRAESTRUCTURA COMO CÓDIGO

Es la práctica de **utilizar scripts** para configurar la infraestructura de computación.

CloudFormation | Deploy Manager

ORQUESTACIÓN DE CARGAS



Después de revisar las diferentes opciones que proveen los cloud providers encontramos variedad en servicios de acuerdo a su funcionalidad, otras nubes como Azure, Softlayer, Alibaba también cuentan con servicios orientados al procesamiento de datos, sin embargo dentro de su ecosistema no es tan completo el set de servicios, por tal motivo siempre que pensemos en proyectos de BigData los mejores cloud provider serán AWS y GCP que estudiaras en este curso.