

Ejercicio:

Arquitectura de datos en la nube

RESUMEN DEL EJERCICIO

El objetivo de este ejercicio es definir una arquitectura simple de una plataforma de datos en una nube donde se almacenarán la información sobre diferentes fuentes sobre los Objetivos del Desarrollo Sostenible para poder pronosticar si con los actuales proyectos en marcha se conseguirán dichos objetivos.

Ejercicio 1. Diseñar una plataforma de datos en la nube y calcular su coste.

Se quiere construir una plataforma de datos en una nube para almacenar y analizar información los Objetivos del Desarrollo Sostenible y si los proyectos en los que se está trabajando van a ayudar a conseguirlos.



Para ello, se ha pedido a un Arquitecto de Datos que diseñe un lago de datos y calcule su coste.

En el proyecto, trabajará un Ingeniero de Datos que se encargará de crear los data pipelines que extraerán la información de las diferentes fuentes de datos y las guardará en su formato

original en el lago de datos, manteniendo un histórico de datos. Los datos vendrán de diferentes fuentes:

- Los indicadores generales e información de los últimos 20 años se obtienen de las APIs del World Bank en CSVs ([Data Catalog \(worldbank.org\)](https://datacatalog.worldbank.org/)) (1 Gb)
- Los datos de los proyectos en los que se trabaja actualmente se obtienen de una BD SQL (5Gb)
- Información climática (temperaturas, lluvias, presión, humedad, etc) de los países a estudiar [Descargar | Servicio de Información Meteorológica Mundial \(wmo.int\)](https://wmo.int) (1Gb)

A parte de tener la información en formato original y manteniendo un histórico, se desea tener una sección con los últimos datos cargados pero limpios para que sean fácilmente consultados por el Científico de Datos.

Finalmente, se tendría un apartado donde tras procesar los datos y formatearlos según el modelo en el que se está trabajando.

El Analista de Datos se encargará de construir un modelo de datos que va cambiando tal y como se añaden nuevas fuentes, al ir identificando nuevas dimensiones y medidas sobre los datos. Se estima que el modelo se recalcula 5 veces al día y se hacen llamadas previas para estudiar y analizar la información sobre 10 veces al día.

El Científico de Datos trabaja parcialmente en este proyecto 6 horas al día, 3 días a la semana construyendo un modelo para predecir si con los proyectos actuales se van a conseguir alcanzar los Objetivos de Desarrollo Sostenible o no.

Para ello, el Científico necesitará un servicio que le ayude a analizar los datos haciendo consultas en memoria y aquellas que sean útiles, las guardaran en una base de datos.

A parte, a los Científicos de Datos también necesitarán un servicio en que pueda aplicar modelos de Machine Learning y generar conjuntos de datos procesados que guardarán en el lago de datos.

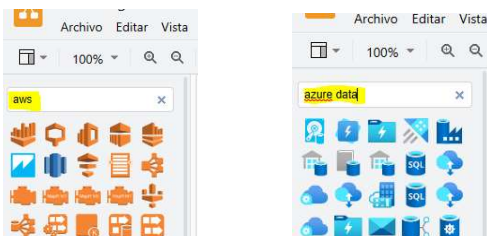
Finalmente, se puede visualizar la información utilizando una herramienta de BI que se conectará directamente la base de datos con los datos útiles y formateados almacenados en el lago de datos.

A ENTREGAR

- Con la información proporcionada en el apartado anterior, construir un **diagrama de la arquitectura** de la plataforma de datos en el cloud para el ejercicio propuesto. Puede realizarse utilizando cualquier herramienta y tan solo hace falta enviar la imagen del resultado final (4 puntos)
- En base al diseño presentado, calcular el posible **coste de la plataforma de datos** de manera mensual y exportar la lista detallada de los gastos por los servicios en un csv. (2 puntos)

Notas:

1. Sugerencia: herramienta gratuita para definir el diagrama de arquitectura de la plataforma de datos [draw.io \(drawio.com\)](https://draw.io) // <https://app.diagrams.net/>



2. Podéis consultar las siguientes webs para poder calcular el coste mensual aproximado de la Plataforma de Datos.
 - AWS: <https://calculator.aws/>
 - Azure: [Calculadora de precios | Microsoft Azure](#)