

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERIA Y ARQUITECTURA
ESCUELA DE INGENIERIA DE SISTEMAS INFORMATICOS
ESPECIALIZACION: INGENIERO DE DATOS

MODULO 3: **PROCESAMIENTO DE DATOS**
PROFESOR: ING. RENE FABRICIO QUINTANILLA

NOTA FINAL

EVALUACION

ORDINARIA



DIFERIDA



No. DE LISTA:

ALUMNO:

CARNET:

FECHA:

INSTRUCCIONES:

- a.- Lea cuidadosamente el examen y esté seguro de que ha comprendido todos y cada uno de los puntos solicitados.
- b.- Para que un punto sea valido en su totalidad, deberá demostrar que ha realizado el ejercicio completo según lo solicitado.
- c.- Sea ordenado en el desarrollo de la prueba, evitando inconsistencias o puntos incompletos **que dejen duda de sus respuestas.**
- d.- Deberá crear un video explicando el paso a paso realizado y entregarlo a más tardar el viernes 27 de octubre 2023 a la media noche a través del aula virtual.

Desarrolle cada uno de los siguientes puntos:

La empresa aeronautica "Para Volar" tiene en operación el sistema de vuelos "Gamatron", el cual se encarga de operativizar los vuelos comerciales realizados por la empresa hacia diferentes aeropuertos en el mundo a través de reservaciones de vuelos y de abordajes por vuelos realizados.

Después de varios años operando, la compañía ha decidido contratarlo para diseñar su estrategia de Big Data, para lo cual le comparte la información que se descarga de sus sistemas donde se muestran datos sobre estados de vuelos y se busca que esta sea cargada en un repositorio donde se pueda exponer dicha información de forma ágil. Esta información se puede descargar de este enlace:

<https://www.kaggle.com/datasets/robikscube/flight-delay-dataset-20182022>

Para comenzar a implementar su estrategia de Big Data, el negocio requiere hacer una proyección de vuelos cancelados, retrasados y diferidos, así como las causas del posible incidente (clima, llantas, seguridad del viaje, llegadas tardías, etc). Para solventar este requerimiento, la empresa le solicita lo siguiente:

1. (25%) Diseñar una base de datos para almacenar el datawarehouse con el que se analizará esta información
2. (20%) Cree una estructura de carpetas para procesar la información en Amazon S3
3. (30%) Cree una base de datos en Amazon Redshift siguiente el diseño de la base de datos elaborada en el punto 1.
4. (25%) Construya procesos ETL para la carga de información en S3 y su procesamiento para prepararlo para su carga en Amazon Redshift.