UIVERSIDAD DE EL SALVADORFACULTAD DE INGENIERIA Y ARQUITECTURAESCUELA DE INGENIERIA DE SISTEMAS INFORMATICOSESPECIALIZACIÓN EN INGENIERÍA DE DATOS



**Resolución parcial #3**

|  |  |  |
| --- | --- | --- |
| **Nombres** | **Apellidos** | **Carnet** |
| **Aquino Alas** | **Doris Elizabeth** | **AA11002** |

**Docente: Ing. René Fabricio Quintanilla**

**Cuidad Universitaria, 29 de octubre de 2023**

ÍNDICE

[Enunciado parcial 3](#_Toc149410687)

[Diagrama del Dataset 4](#_Toc149410688)

[Diccionario de datos 5](#_Toc149410689)

[Modelo dimensional propuesto 6](#_Toc149410690)

[Pasos para el modelado dimensional 6](#_Toc149410691)

[Diagrama dimensional 8](#_Toc149410692)

[Mapping de tablas 9](#_Toc149410693)

# Enunciado parcial

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.

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.

# 

# Diagrama del Dataset

Tabla

Descripción generada automáticamente con confianza media

# Diccionario de datos

**Tabla**: Airlines.

**Descripción**: Contiene datos generales de las aerolíneas.

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Nombre | Tipo de dato | Descripción |
| 1 | Code | String | Código único identificador de la aerolínea. |
| 2 | Description | String | Nombre de la aerolínea. |

**Tabla**: Flights

**Descripción**: Contiene los detalles del vuelo además de cancelaciones y retrasos.

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Nombre | Tipo de dato | Descripción |
| 1 | FlightDate | String | Fecha de vuelo |
| 2 | Flight\_Number\_Operating\_Airline | Integer | Número de vuelo |
| 3 | Origin | String | Abreviación del aeropuerto de salida |
| 4 | OriginCityName | String | Nombre de la ciudad de salida |
| 5 | OriginStateName | String | Nombre del estado de salida |
| 6 | Dest | String | Abreviación del aeropuerto de llegada |
| 7 | DestCityName | String | Nombre de la ciudad de llegada |
| 8 | DestStateName | String | Nombre del estado de llegada |
| 9 | CRSDepTime | Integer | Hora programada de salida |
| 10 | DepTime | String | Hora real de salida |
| 11 | DepDelayMinutes | String | Minutos de retraso en salida |
| 12 | CRSArrTime | Integer | Hora programada de llegada |
| 13 | ArrTime | String | Hora real de llegada |
| 14 | ArrDelayMinutes | String | Minutos de retraso en llegada |
| 15 | WheelsOff | String | Hora que el avión se levanta de la tierra (Salida) |
| 16 | WheelsOn | String | Hora que el avión toca tierra (Llegada) |
| 17 | Cancelled | Integer | Indicador de vuelo cancelado |
| 18 | Diverted | Integer | Indicador de vuelo desviado |
| 19 | AirTime | String | Tiempo de vuelo en el aire |
| 20 | Flights | Integer | Cantidad de vuelos |
| 21 | CarrierDelay | String | Minutos de retraso debido al transportador |
| 22 | WeatherDelay | String | Minutos de retraso debido al clima |
| 23 | NASDelay | String | Minutos de retraso en el Sistema Nacional de Aero espacio |
| 24 | SecurityDelay | String | Minutos de retraso debido a seguridad |
| 25 | LateAircraftDelay | String | Minutos de retraso de la aeronave |
| 26 | DivReachedDest | String | Indicador si un vuelo desviado, logró alcanzar su destino inicial programado. |

# Modelo dimensional propuesto

## Pasos para el modelado dimensional

1. **Proceso:** Análisis de vuelos
2. **Granularidad:** Una fila de datos representa un vuelo diferido, cancelado o atrasado por diferentes casusas clima, llantas, transportador, NAS, seguridad, retraso de la aeronave, filtrado por fecha, ciudad de origen y destino y por aerolínea.
3. **Dimensiones:**

DimAirline

* AirlineId
* AirlineKey
* Description

DimDate

* DateKey
* fullDate
* dayName
* datOfWeek
* dayNumInMonth
* monthName
* monthNumOverall
* weekNumInYear
* quarterName
* quarter
* semesterName
* semester
* year

1. **Métricas**

* **Cantidad de vuelos retrasados** ->por fecha, ciudad de origen y destino, por aerolínea,
* **Cantidad de vuelos diferidos** ->por fecha, ciudad de origen y destino, por aerolínea,
* **Cantidad de vuelos cancelados** ->por fecha, ciudad de origen y destino, por aerolínea,
* **Cantidad de vuelos retrasados según el motivo** (clima, llantas, transportador, NAS, seguridad, retraso de la aeronave) ->por fecha, ciudad de origen y destino, por aerolínea
* **Total de tiempo (en minutos) de retrasos según el motivo** (clima, llantas, transportador, NAS, seguridad, retraso de la aeronave) ->por fecha, ciudad de origen y destino, por aerolínea
* **Porcentaje de vuelos cancelados** ((Total de vuelos programados/100) \*vuelos cancelados)
* **Tiempo total de vuelo por fecha y aerolínea**

FactFlights

* AirlineKey
* DateKey
* Flight\_Number\_Operating\_Airline
* Origin
* OriginCityName
* OriginStateName
* Dest
* DestCityName
* DestStateName
* Flights
* AirTime
* Cancelled
* Diverted
* Delayed
* WheelsOffMinutes
* WheelsOnMinutes
* DepDelayMinutes
* ArrDelayMinutes
* CarrierDelay
* WeatherDelay
* NASDelay
* SecurityDelay
* LateAircraftDelay
* TotalDelayTime
* DivReachedDest

## Diagrama dimensional

Tabla

Descripción generada automáticamente

# 

# Mapping de tablas

Para efectos de poder visualizar mejor el mapping de tablas debe dirigirse al siguiente enlace de GitHub donde se encuenta el archivo Mapping\_Tablas.xlms:

<https://github.com/aquino08/proyecto_parcial3_aa11002/tree/main/Mapping_Tablas>