**UIVERSIDAD DE EL SALVADOR**

**FACULTAD DE INGENIERIA Y ARQUITECTURA**

**ESCUELA DE INGENIERIA DE SISTEMAS INFORMATICOS**

**ESPECIALIZACIÓN EN INGENIERÍA DE DATOS**



**Parcial N.º 3**

**Catedrático:** Ing. René Fabricio Quintanilla Gómez.

**Estudiante:** Eduardo José Orellana Guerretta.

**Carnet:** OG17006.

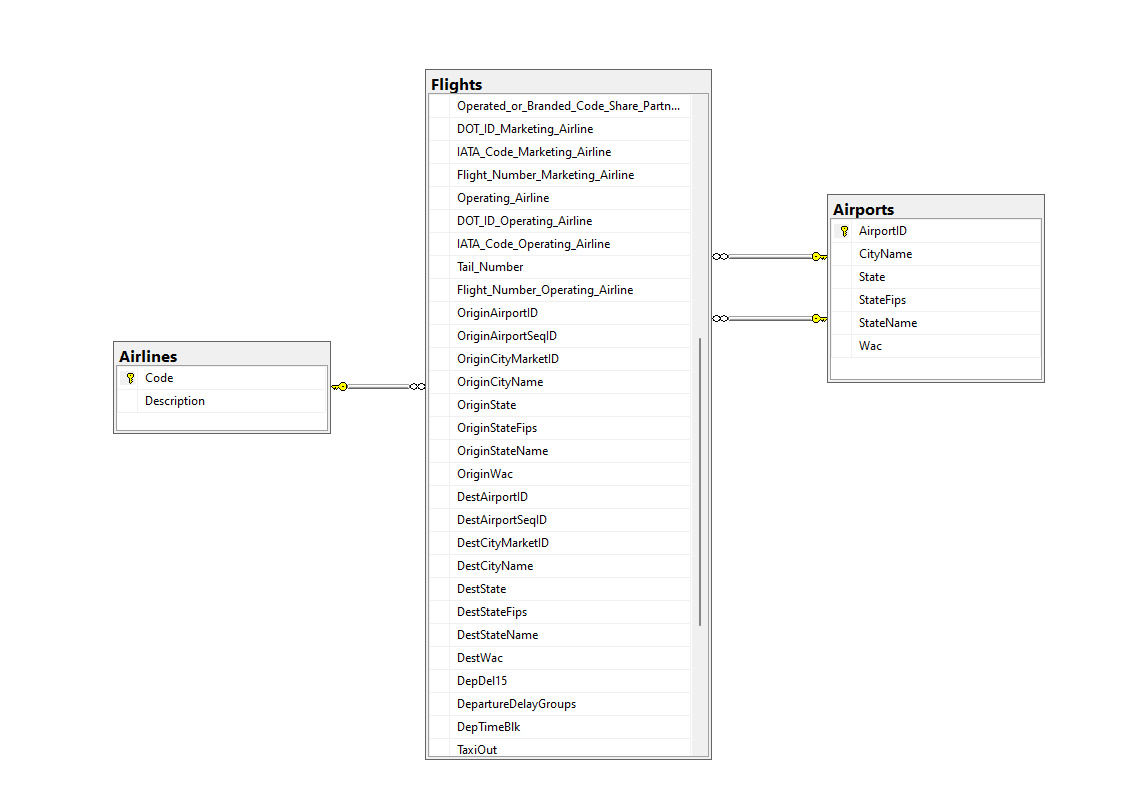
Ciudad Universitaria, San Salvador, 25 de octubre de 2024.

**Introducción**

En el presente documento, se presenta el análisis llevado a cabo para poder construir una solución de Big Data para la construcción de DW que le permita a la empresa “Para Valor”, poder analizar los datos referentes a vuelos que pudieron haber sido cancelados, retrasados o diferidos.

**Diagrama del Dataset**

El Data set únicamente está compuesto por la tabla de las aerolíneas y de los vuelos (originalmente en formato .csv por cada año), pero, el archivo que posee los vuelos, tenia la información de los aeropuertos de origen y destino, por ello, se separa la info de los aeropuertos en otra tabla aparte



**Modelo Dimensional Propuesto**

Pasos para la creación del modelo dimensional:

1. Selección del proceso: Análisis de Vuelos.
2. Granularidad: Una fila de datos, representa un vuelo realizado en **“X”** fecha de un Aeropuerto **“A”**, a un Aeropuerto **“B”** por una Aerolínea.
3. Dimensiones:
   1. DimTiempo

* TiempoKey
* Fecha
* Dia
* Mes
* Anio
* Trimestre
* DiaDeLaSemana
  1. DimAerolinea
* AerolineaKey
* CodigoAerolinea
* NombreAerolinea
* FechaDeActualizacion (SCD2)
* Activo (SCD2)
  1. DimAeropuerto
* AeropuertoKey
* AeropuertoId
* NombreCiudad
* AbreviaturaEstado
* FipsEstado (Federal Information Processing Standards)
* NombreEstado
* CodigoAreaMundial

1. Métricas:

Métricas Aditivas

* **Distancia**: La distancia total recorrida por el vuelo en millas.
* **Tiempo de vuelo**: El tiempo total que el vuelo pasó en el aire desde el despegue hasta el aterrizaje en minutos.
* **Tiempo en aire**: El tiempo efectivo que el avión estuvo en vuelo sin incluir tiempos de taxi.
* **Tiempo transcurrido real**: El tiempo total que tomó el vuelo desde el despegue hasta el aterrizaje, incluyendo cualquier retraso o desvío.
* **Aterrizajes desviados**: El número total de aterrizajes que fueron desviados a un aeropuerto diferente del planificado.
* **Diferencia de tiempo de salida**: La diferencia en minutos entre la hora prevista de salida y la hora real de salida.
* **Diferencia de tiempo de llegada**: La diferencia en minutos entre la hora prevista de llegada y la hora real de llegada.
* **Tiempo en ruta a la puerta**: El tiempo en minutos que el avión tardó en moverse desde la pista de aterrizaje hasta la puerta de desembarque (taxi in).
* **Tiempo en ruta a la pista**: El tiempo en minutos que el avión tardó en moverse desde la puerta de embarque hasta la pista de despegue (taxi out).

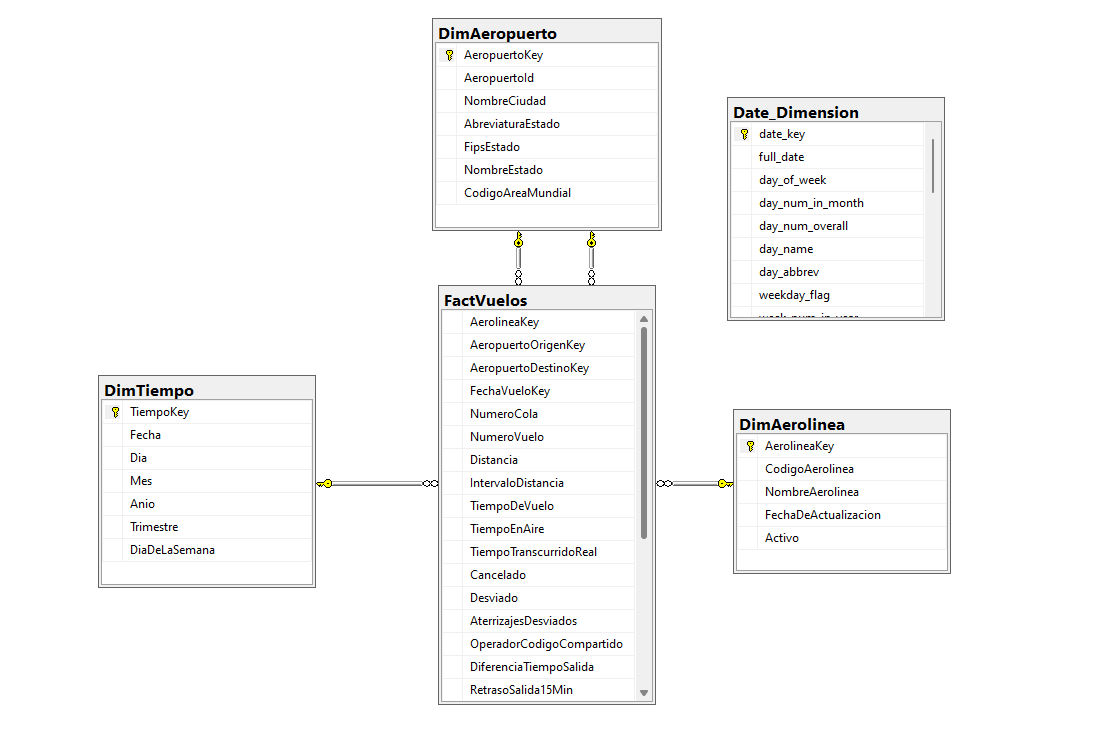
Métricas Semi Aditivas

* **Intervalo de distancia**: El intervalo de distancia en millas en el que cae el vuelo, generalmente agrupado en rangos.
* **Intervalo de retraso en salida**: El intervalo de tiempo en minutos que agrupa el retraso de la salida del vuelo.
* **Intervalo de retraso en llegada**: El intervalo de tiempo en minutos que agrupa el retraso de la llegada del vuelo.

Métricas No Aditivas

* **Cancelado**: Indica si el vuelo fue cancelado o no (Verdadero/Falso).
* **Desviado**: Indica si el vuelo fue desviado a otro aeropuerto o no (Verdadero/Falso).
* **Operador en código compartido**: El código del operador con el que se comparte el vuelo bajo un acuerdo de código compartido.
* **Retraso en salida de más de 15 minutos**: Indica si el vuelo tuvo un retraso mayor o menor a 15 minutos en la salida (Retraso/Sin Retraso).
* **Retraso en llegada de más de 15 minutos**: Indica si el vuelo tuvo un retraso mayor o menor a 15 minutos en la llegada (Retraso/Sin Retraso).
* **Hora programada de salida**: La hora específica en la que estaba programada la salida del vuelo (formato "hh:mm").
* **Hora programada de llegada**: La hora específica en la que estaba programada la llegada del vuelo (formato "hh:mm").
* **Intervalo de hora de salida**: El rango de tiempo en el que está programada la salida del vuelo (ej. "hh:mm-hh:mm").
* **Intervalo de hora de llegada**: El rango de tiempo en el que está programada la llegada del vuelo (ej. "hh:mm-hh:mm").
  1. FactVuelos:
* AerolineaKey
* AeropuertoOrigenKey
* AeropuertoDestinoKey
* FechaVueloKey
* NumeroCola
* NumeroVuelo
* Distancia
* IntervaloDistancia
* TiempoDeVuelo
* TiempoEnAire
* TiempoTranscurridoReal
* Cancelado
* Desviado
* AterrizajesDesviados
* OperadorCodigoCompartido
* DiferenciaTiempoSalida
* RetrasoSalida15Min
* IntervaloRetrasoSalida
* DiferenciaTiempoLlegada
* RetrasoLlegada15Min
* IntervaloRetrasoLlegada
* TiempoEnRutaALaPuerta
* TiempoEnRutaALaPista
* IntervaloHoraSalida
* IntervaloHotaLlegada
* HoraProgramadaSalida
* HoraProgramadaLlegada

**Diagrama Dimensional para el Proceso: Análisis de Vuelos**



**Script de Creación de Modelo Dimensional**

Los scripts a utilizar para el modelo dimensional, se encuentran dentro de la carpeta SQL, cada archivo está nombrado en base a su función.

**Anexos**

Las imágenes del modelo relacional y dimensional se encuentran dentro de la carpeta **Images**.