

CASO PRACTICO - 1 - diegoserranoventurini

NOTA INICIAL: La descripción de los datos proporcionados se realiza en el punto 3 del ejercicio; sin embargo, para dar una respuesta completa a las preguntas planteadas en el resto de apartados es imprescindible haber analizado previamente los datos. Por ello, el hilo argumental de las respuestas queda un tanto confuso ya que se asume un conocimiento de los datos que formalmente no ha sido planteado.

Análisis de requerimientos

En este apartado se pretende realizar un estudio de los posibles casos de uso y necesidades que la construcción de este DWHU (*Data Warehouse Unified*) pretende dar respuesta.

Los agentes que se especifican en el enunciado son varios, y de naturaleza e intereses muy diversos (y contrapuestos en algunos casos). Los que se enumeran, sin detallar sus intereses particulares son:

- Sector turismo: OMT, agencias de viajes, cadenas hoteleras...
- Aerolíneas
- Industria aerocomercial: tiendas en los aeropuertos, empresas de servicios alrededor de aeropuertos, servicios auxiliares aeroportuarios, ...
- Administración pública
- Ciudadanos

En este caso práctico/proyecto se pretende construir un DWHU que centraliza cierta información relevante a la actividad económica generada por las aerolíneas y el tráfico de viajeros en los aeropuertos. El principal uso de esta información es conocer la evolución del número de pasajeros entre países; la actividad de las aerolíneas; las rutas entre los países y el impacto que estos números han tenido en sus ingresos vinculados al turismo.

Podemos identificar por tanto las siguientes necesidades para varios de los grupos interesados:

- Sin duda tanto para las administraciones públicas como para las privadas interesante conocer la evolución del impacto del turismo en las economías nacionales. Encontramos que existen varias métricas socio-económicas interesantes en el conjunto de datos. Si un país está experimentando una tendencia ascendente de visitantes y de ingresos puede plantearse inversiones en infraestructuras; mientras que otro que vea su número de visitantes disminuir puede decidir aumentar su inversión en marketing y posicionamiento.
- Para el sector turismo privado es importante conocer la actividad económica de las diferentes aerolíneas, donde operan, con qué equipamiento y con qué volumen. Por ejemplo, una empresa que proporcione servicios de

mantenimiento de aviones puede detectar lugares en los que sus *partners* (aerolíneas) estén operando y ellos no.

- Una necesidad que satisficiera este DWH sería conocer el grado de conexión de los diferentes países, es decir, cuánto tráfico aéreo tienen de salida y entrada. Esta métrica ayudaría a saber si un país está más aislado o si tiene una mayor exposición al turismo. Esto abriría la puerta a que las aerolíneas puedan plantear nuevas rutas a países pobremente conectados.
- Finalmente los ciudadanos conocerían el grado de expansión de las aerolíneas y los códigos de todas las rutas y la actividad económica que la actividad de estas genera. De esta manera, tendrían más argumentos para poder cuestionar políticas orientadas al turismo; y un mayor conocimiento de los lugares que podrían emplear como destino vacacional.

En resumen, este DWH permitiría conocer la evolución macro de la actividad turística de los países y cuán abiertos están al resto del mundo; así como el detalle de la actividad de las aerolíneas en estos países.

Indicadores, hechos, dimensiones, atributos

A continuación, como parte del proceso de definición del DWHU vamos a identificar los hechos y dimensiones con sus métricas y sus atributos dentro de los datos proporcionados. Este contenido se empleará y extenderá con los hechos derivados en el diseño formal del modelo multidimensional.

Las dimensiones identificadas son las siguientes:

- **Dimensión Lugar.** Existen tres niveles en esta dimensión: el aeropuerto, la ciudad y el país. Cada uno de los cuales tiene varios atributos específicos. Algunos ejemplos son: geoposicionamiento, código identificativo, uso horario, continente...
- **Dimensión Tiempo.** La granularidad de los datos es anual, ya que no se proporcionan hechos en otra unidad: días, horas, o meses, etc.
- **Dimensión Aerolínea.** Aunque existen aerolíneas que son secciones de otras (como las low-cost de compañías grandes) no se especifica dicha relación de jerarquía por lo que solo existe un nivel de granularidad.
- **Dimensión Indicadores.** Esta dimensión ayuda a identificar los hechos en las tablas históricas.
- **Dimensión Equipamiento.** Esta dimensión indica el tipo de tecnología o aeronave empleada en el desempeño de la actividad.

En cuanto a hechos encontramos los siguientes:

- **Rutas.** Este hecho podría considerarse casi dimensión: es una intersección entre varias dimensiones anteriores (Lugar - Aerolínea); contiene atributos propios como **stops** (número de paradas), e interpretamos que es un “punto de vista desde el que se pueden analizar los datos”, ya que se podrían generar reportes como número de rutas que un país contiene.

de entrada. Dicha metrica daria respuesta a uno de los requerimientos planteados.

- Numero de pasajeros anuales.
- Total de ingresos generados por el turismo en un pais.
- Porcentaje del GDP dedicado al intercambio comercial
- Numero de aerolineas que operan en un pais
- Numero de conexiones entre dos paises
- Numero de rutas entrantes/salientes de un pais

Como se ha comentado en el apartado del disenho multidimensional se indicara como se agrupan dichas dimensiones y hechos en el DWHU; asi como el detalle de los atributos existentes.

Analisis del *dataset* proporcionado

En este apartado analizaremos todas las tablas proporcionadas, indicando para cada tabla los campos que contienen, su nombre (original y castellano), una breve descripcion del dato, su tipo y el nivel de informacion que contiene.

airlines.dat

Atributo	Atributo (ESP)	Desc	Tipo	Nivel inf
Airline ID	ID aerolinea	Identificador unico de la aerolinea en OpenFlights	INT	1 nivel con 6048 tipos
Name	Nombre	Nombre de la aerolinea	STR	1 nivel
Alias	Alias	Alias por el que se conoce la aerolinea	STR	1 nivel
IATA	IATA	IATA codigo de 2 letras (si esta disponible)	STR	1 nivel
ICAO	ICAO	ICAO codigo de 3 letras (si esta disponible)	STR	1 nivel
Callsign	Senal	Nombre por el que se llama a las aerolineas	STR	1 nivel
Country	Pais	Pais o territorio donde reside la aerolinea	STR	1 nivel

Atributo	Atributo (ESP)	Desc	Tipo	Nivel inf
Active	Activa	Flag que indica si la aerolinea esta activa o no	STR	1 nivel (equivalente a Y/N)

airports.dat

Atributo	Atributo (ESP)	Desc	Tipo	Nivel inf
----------	----------------	------	------	-----------

Airport ID ID aeropuerto Identificador unico INT 1 nivel del aeropuerto en OpenFlights

Name Nombre Nombre del STR 1 nivel aeropuerto

City Ciudad Nombre de la STR 1 nivel ciudad donde se encuentra el aeropuerto

Country Pais Nombre del STR 1 nivel pais donde se encuentra el aeropuerto

IATA IATA IATA codigo de STR 1 nivel 3 letras

ICAO ICAO ICAO codigo de STR 1 nivel 4 letras

Latitude Latitud Grados decimales FLOAT 6 (negativo es sur, decimales positivo es norte)

Longitude Longitud Grados decimales FLOAT 6 (negativo es decimales oeste, positivo es este)

Altitude Altitud Altitud en FLOAT 6 "pies" decimales

Timezone Uso horario Horas de diferencia FLOAT 1 decimal respecto a UTC

DST DST Horario de STR 1 nivel - invierno (7 valores E - EUROPA, A - US/CAN, S - America Sur 0 - AUST Z - Nueva Zel. N - None U - Desconocido)

TZ Zona horaria Zona horaria en STR 1 nivel Olson

Type Tipo Tipo aeropuerto STR 1 nivel - 4 valores

routes.dat

Atributo	Atributo (ESP)	Desc	Tipo	Nivel inf
Airline ID	ID aerolinea	Identificador unico de la aerolinea en OpenFlights	INT	1 nivel con 6048 tipos
Airline	Nombre aerolinea	Nombre de la aerolinea	STR	1 nivel
Source airport	Aeropuerto origen	codigo IATA/ICAO del aeropuerto de origen	STR	1 nivel
Source airport ID	ID Aeropuerto origen	codigo OpenFlights del aeropuerto de origen	STR	1 nivel
Destination airport	Aeropuerto destino	codigo IATA/ICAO del aeropuerto de destino	STR	1 nivel
Destinatio airport ID	ID Aeropuerto destino	codigo OpenFlights del aeropuerto de destino	STR	1 nivel
Codeshare	Compartido	indicador de si el vuelo esta compartido	STR	1 nivel
Stops Equipment	Paradas Equipamiento	numero de paradas tres letras indicando el equipamiento	INT STR	1 nivel

worldbank_comercio_ext_anual_pais.csv

Atributo	Atributo (ESP)	Desc	Tipo	Nivel inf
Country Name	Nombre Pais	Nombre del pais	STR	1 nivel
Country Code	Codigo Pais	Codigo del pais	STR	1 nivel
Indicator Code	Codigo Indicador	Codigo del indicador	STR	1 nivel
Indicator Name	Nombre Indicador	Nombre del indicador	STR	1 nivel
YEAR	ANHO	Sucesion de anhos	INT	anhos desde 1960

worldbank_ing_turismo_anual_pais.csv

Atributo	Atributo (ESP)	Desc	Tipo	Nivel inf
Country Name	Nombre Pais	Nombre del pais	STR	1 nivel
Country Code	Codigo Pais	Codigo del pais	STR	1 nivel
Indicator Code	Codigo Indicador	Codigo del indicador	STR	1 nivel
Indicator Name	Nombre Indicador	Nombre del indicador	STR	1 nivel
YEAR	ANHO	Sucesion de anhos	INT	anhos desde 1960

worldbank_pasajeros_anual_pais.csv

Atributo	Atributo (ESP)	Desc	Tipo	Nivel inf
Country	Codigo Pais	Codigo del pais	STR	1 nivel
Indicator	Codigo Indicador	Codigo del indicador	STR	1 nivel
Indicator	Nombre Indicador	Nombre del indicador	STR	1 nivel
YEAR	ANHO	Sucesion de anhos	INT	anhos desde 1960

worldbank_poblacion_anual_pais.csv

Atributo	Atributo (ESP)	Desc	Tipo	Nivel inf
Country Code	Codigo Pais	Codigo del pais	STR	1 nivel
Indicator Code	Codigo Indicador	Codigo del indicador	STR	1 nivel
Indicator Name	Nombre Indicador	Nombre del indicador	STR	1 nivel
YEAR	ANHO	Sucesion de anhos	INT	anhos desde 1960

equipamientos.js

Atributo	Atributo (ESP)	Desc	Tipo	Nivel inf
desc_equipamiento	desc_equipamiento	Descripcion equipamiento (aeronave)	STR	1 nivel
cod_equipamiento	cod_equipamiento	Codigo equipamiento	STR	1 nivel

países.xml

Atributo	Atributo (ESP)	Desc	Tipo	Nivel inf
cod_pais	cod_pais	Codigo pais	STR	1 nivel
cod_pais2	cod_pais2	Codigo alternativo pais	STR	1 nivel
desc_pais	desc_pais	Nombre pais	STR	1 nivel
cod_continente	cod_continente	codigo continente	STR	1 nivel
desc_continente	desc_continente	nombre continente	STR	1 nivel
latitute	Latitud	Grados decimales	FLOAT 6 decimales	(negativo es sur, positivo es norte)
longitude	Longitud	Grados decimales	FLOAT 6 decimales	(negativo es oeste, positivo es este)

Analisis funcional

A continuacion se van a especificar los requisitos funcionales solicitados en la contruccion de este DWHU.

Numero	Requerimiento	Prioridad	Exigible
1	Se cargara la informacion proporcionada en el sistema, cuidando de solo cargar aquellos datos relevantes	1	SI
2	Se creara un DWHU para satisfacer las necesidades de todos los agentes especificados	1	SI
3	Se creara un modelo OLAP para la consulta de la informacion multidimensional	2	SI
4	Se crearan vistas OLAP realizando varios analisis requeridos	3	NO

Diseño del modelo conceptual, lógico y físico

Realizado el analisis de los datos y de los requerimientos pasamos a analizar que como se va a construir el DWHU unificado, que tablas, dimensiones, atributos, etc.

Diseño conceptual

Analizando la informacion de la que disponemos encontramos dos tablas de hechos con varias dimensiones cada una que se especificaran a continuacion.

Tabla de hecho	Descripcion
h_KPI_socio_economico	Incluye resultados socio economicos para un LUGAR, TIEMPO y un INDICADOR
h_actividad_aeropuerto	Describe la operativa de las aerolineas en el desarrollo de su actividad

En la primera tabla encontramos tres dimensiones

- LUGAR: con dos niveles `d_l_pais` y `d_l_continente`.
- TIEMPO: con un nivel `d_t_anho`. Esta dimension podria ampliarse a `d_t_decada`
- INDICADOR: con un nivel `d_i_indicador`.

y una metrica:

- VALOR: con el valor numerico del INDICADOR para ese LUGAR y TIEMPO.

En la segunda tablas encontramos las dimensiones

- AEROLINEA: con un nivel `d_a_aerolinea`
- LUGAR: con cuatro niveles `d_l_aeropuerto`, `d_l_ciudad`, `d_l_pais` y `d_l_continente`.
- EQUIPAMIENTO: con un nivel `d_e equipamiento`

El hecho seria la ruta existente interseccion de las dimensiones anteriores, con atributos:

- PARADAS
- COMPARTIDO

Diseño físico

A continuacion vamos a implementar el disenho conceptual anterior en Oracle.

Primero comenzamos con las tablas de dimensiones:

d_lugar_aeropuerto

Esta tabla tiene los siguientes campos y es la dimension mas compleja que encontramos:

- id_aeropuerto: clave primaria. numero
- cod_aeropuerto: cod aeropuerto. varchar(3)
- cod_aeropuerto2: cod aeropuerto. varchar(4)
- cod_ciudad: cod ciudad. varchar(22)
- desc_ciudad: desc ciudad. varchar2(22)
- id_pais: codigo pais. varchar2(22)
- desc_pais: desc pais. varchar2(22)
- id_continente. codigo pais. varchar2(22)
- desc_continente. desc continente. varchar2(22)
- latitud: numero
- longitud: numero
- altitud: numero
- zona horaria: varchar(22)

d_lugar_pais

Esta tabla tiene los siguientes campos y es la dimension mas compleja que encontramos:

- id_pais: clave primaria pais. varchar2(22)
- desc_pais: desc pais. varchar2(22)
- id_continente. codigo pais. varchar2(22)
- desc_continente. desc continente. varchar2(22)
- latitud_pais: numero
- longitud_pais: numero

d_tiempo

Esta tabla tiene los siguientes campos:

- year: clave primaria. number
- decada: number

d_indicador

Esta tabla tiene los siguientes campos:

- id_indicador: clave primaria. number
- cod_indicador: codigo indicador. varchar(22)
- desc_indicador: desc indicador. varchar(50)

d_aerolinea

Esta tabla tiene los siguientes campos:

- id_aerolinea: clave primaria. number
- cod_aerolinea: codigo aerolinea. varchar(2)
- cod_aerolinea: codigo aerolinea 2. varchar(3)
- desc_aerolinea: desc aerolinea. varchar(50)
- id_pais. clave foranea. number

d equipamiento

Esta tabla tiene los siguientes campos

- id_equipamiento. clave primaria. number
 - cod_equipamiento. varchar(22)
 - desc_equipamiento. varchar2(255)
-

Finalmente las tablas de hechos:

h_KPI_socio_economico

Contiene los campos:

- year
- id_pais
- id_indicador
- value

h_actividad_aeroportuaria

Contiene los campos

- id_aerolinea
- id_aeropuerto1
- id_aeropuerto2
- equipamiento
- paradas
- compartida