
Temática: *Data Warehouse* / Almacenes de datos

Caso de estudio: Diseño de un almacén de datos para la gestión de información de aeropuertos, tráfico aéreo e indicadores turísticos.

Autor: Juan Vidal Gil

Revisado por: Josep Curto Díaz

Índice

- Introducción
- Contexto
- Programas
- Bibliografía

Introducción

El contexto del material que presentamos a continuación está basado en el uso de datos reales obtenidos de la sección de *open data* del banco mundial y de bases de datos de aeropuertos disponibles también en iniciativas *open data*.

Este material, titulado “Diseño de un almacén de datos para la gestión de información de aeropuertos, tráfico aéreo e indicadores turísticos”, está diseñado para practicar el diseño e implementación del núcleo de toda herramienta de inteligencia de negocio: el almacén de datos.

Consta de tres partes diferenciadas. La primera, tiene por objetivo desarrollar e implementar un almacén de datos que permita la gestión de la información de aeropuertos, tráfico aéreo e indicadores relacionados; la segunda, se centra en diseñar e implementar los procesos de carga de datos necesarios para disponer de información en almacén de datos implementado; y la tercera en la creación de elementos de análisis multidimensional para la explotación de información.

Con el fin de poder desarrollar un proyecto lo más específico posible, el estudiante tendrá que afrontar el reto de desarrollar un almacén de datos que sólo describe parte de los servicios que se pueden ofrecer en base a datos de aeropuertos y tráfico aéreo que formarían parte de un sistema real.

Es importante recordar que el diseño, desarrollo e implantación de este tipo de herramientas supone, en general meses o incluso años en función de la naturaleza y el grado de madurez de la empresa y del alcance del proyecto, la participación de equipos multidisciplinares que van implementando diferentes proyectos en un proceso de mejora continua.

En este sentido, el objetivo de este caso no es desarrollar un almacén de datos que dé respuesta a todas las necesidades de la organización, sino entender y utilizar las metodologías para desarrollar este tipo de proyectos en un contexto real pasando por todas las fases que comprenden este tipo de proyecto. A saber: evaluación, diseño, implementación, carga y explotación.

A partir de unas necesidades de negocio acotadas, el estudiante deberá adquirir un conocimiento básico del entorno tecnológico, de los procesos de negocio, de las necesidades existentes y definir una propuesta adecuada que responda a ellas.

Mediante el desarrollo del caso, el estudiante se va a encontrar con todos los problemas, dudas y dificultades que se plantean en un proyecto de estas características.

Contexto

El desarrollo de este caso se centrará fundamentalmente en la gestión y reutilización de datos de aeropuertos, tráfico aéreo e indicadores relacionados por parte de una organización especializada en crear valor a partir de la personalización de informaciones existentes en diversos subsectores de la industria, comercio, turismo, etc.

El propósito principal es ofrecer: informes predefinidos o personalizados, cuadros de mando específicos y acceso al análisis libre de datos.

1. Potenciales usuarios del almacén de datos

Existen múltiples potenciales usuarios de datos de aeropuertos, tráfico aéreo e indicadores relacionados. Como punto de partida de cualquier proyecto es necesario entender quiénes son los potenciales usuarios del almacén de datos y a qué tipo de preguntas deberá responder nuestro sistema.

En estos materiales revisaremos primero los diversos actores que podrían beneficiarse de esta solución (siempre y cuando fuera adaptada a sus necesidades) para después introducir el usuario de negocio en el que centraremos todo el caso.

Los actores involucrados en un sistema de datos de tráfico e infraestructura aérea suelen tener intereses muy diversos. Son actores cualquier entidad que se pueda beneficiar de un entorno de análisis de infraestructura aeroportuaria y de métricas de número de pasajeros y otros indicadores relacionados.

En nuestro caso particular, algunos de los actores a tener en cuenta son:

- a) **Sector turismo:** Agentes relacionados con el sector turístico:
 - Ministerios de Turismo
 - Organización Mundial de Turismo
 - Agencias de viajes
 - Cadenas hoteleras

- Secretarías de comercio internacional de los Ministerios de Economía y Competitividad

- b) **Aerolíneas:** las empresas aerolíneas pueden utilizar los datos como apoyo para detectar nuevas posibles rutas, mejora la gestión de la demanda de pasajeros, la ocupación de los vuelos, la estacionalidad, el mercado potencial, identificar destinos turísticos con baja cobertura, optimizar número y tipo de aeronaves, necesidades de personal, etc.

- c) **Industria aerocomercial:** Diferentes agentes relacionados con esta industria como pueden ser:
 - Tiendas *retail* aeropuertos
 - Empresas gestoras de espacios de *parking*
 - Servicios de transporte relacionados con desplazamiento a aeropuertos
 - Empresas que comercialicen soluciones tecnológicas para aeropuertos o aerolíneas
 - Empresas que distribuyan combustible para aeronaves
 - Servicios auxiliares aeropuertos (seguridad, limpieza, etc.)
 - Estudios de investigación universitarios

- d) **Administración pública:** Diferentes ministerios y agencias gubernamentales como pueden ser:
 - Ministerio de Transporte (del país o países vinculantes a los datos)
 - Ministerio de Turismo
 - Ministerio de Energía
 - Agencia de control aéreo
 - Agencia de control de aeropuertos a nivel nacional e internacional

- e) **Ciudadanos:** finalmente los ciudadanos pueden estar interesados en tener una mayor transparencia respecto aerolíneas, aeropuertos, aviones y su servicio.

Está claro que existen múltiples actores interesados y cuyas necesidades van afectar al diseño del almacén de datos.

2. Fuentes de datos

Para el presente caso, cuyo objetivo es el diseño de un almacén de datos y la consecuente carga de información, se proporcionan datos de aeropuertos, aerolíneas y rutas. Datos ofrecidos en abierto en la web: [openflights](https://openflights.org/)

Las iniciativas datos abiertos (*open data*, en inglés) persiguen que determinados tipos de datos estén accesibles de forma libre para todo el mundo, sin restricciones de derechos de autor, de patentes o de otros mecanismos de control. Encontramos con frecuencia datos de este tipo en organizaciones públicas y han tenido un desarrollo considerable en los últimos años, constituyendo una fuente de datos valiosa para los proyectos de analítica o inteligencia de negocio.

Esta información se complementa con datos del [banco mundial](https://data.worldbank.org/). El Banco Mundial es una organización internacional especializada en finanzas que depende de las [Naciones Unidas](https://www.un.org/). En esta web ofrece conjuntos de datos del Banco Mundial de acceso público, así como informes, gráficos y otros recursos. Encontramos en la web multitud de indicadores relacionados con el nivel de desarrollo de los países.

Utilizaremos los datos listados a continuación que están agregados a nivel de año y país y hacen referencia a los siguientes indicadores:

- Movimiento pasajeros anual
- Población países
- Ingresos por turismo
- Datos de comercio exterior

Listado de ficheros:

Nombre fichero	Descripción	Fuente
airpots.dat	Listado de aeropuertos del mundo con conjunto de características (código IATA ¹ , ICAO ² , ciudad, país, coordenadas, zona horaria, etc.)	openflights
airlines.dat	Listado de aerolíneas del mundo con conjunto de características (identificador, código IATA, ICAO, país, atributos actividad, etc.)	openflights
routes.dat	Listado de rutas aéreas del mundo con conjunto de características (aerolínea, aeropuerto origen y destino, paradas, equipamiento,	openflights

¹ International Air Transport Association

² International Civil Aviation Organization

	etc.)	
Equipamientos.js	Aviones de rutas aéreas con código IATA y descripción	aircraft type codes y IATA aircrafts codes
Paises.xml	Listado de países con códigos estándar. Códigos ISO (<i>International Organization for Standardisation</i>) de 2 y 3 caracteres y coordenadas país	Country codes , listado country codes wikipedia y Public data Google Countries
worldbank_pasajeros_anual_pais.csv	Serie histórica de número de pasajeros en vuelos nacionales e internacionales por país y año	Datos pasajeros banco mundial
worldbank_poblacion_anual_pais.csv	Serie histórica de poblaciones a nivel de país y año	Datos poblaciones banco mundial
worldbank_ing_turimo_anual_pais.csv	Serie histórica de ingresos por turismo por país y año (importe en dólares USA)	Datos turismo banco mundial
worldbank_comercio_ext_anual_pais.csv	Serie histórica de datos comercio exterior (suma de exportaciones e importaciones de bienes y servicios representadas como porcentaje del PIB) por país y año	Datos comercio exterior banco mundial

Los nombres de los campos de cada fichero aparecen en la cabecera, salvo los ficheros procedentes de OpenFlights que no tienen cabecera, pero cuya descripción de cambios puede encontrarse en la propia web.

Programas

Para el presente caso, la UOC proporciona una imagen virtual alojada en *Amazon Web Services* (AWS) con las siguientes características:

- Memoria RAM: 2GB.
- Sistema operativo: Windows Server 2012
- Espacio: 60 GB
- Base de datos: Oracle XE 11g
- Herramienta MOLAP: Microsoft SQL Server AnalysisServices 2016
- Herramienta de diseño de ETLs: Spoon – Pentaho Data Integration 7.1
- Herramienta de diseño para AnalysisServices: Visual Studio 2017

Bibliografía

Material de la asignatura Data Warehouse de la UOC.

Kimball, R. (2013) The Data Warehouse Toolkit. Third Edition New York: John Wiley & Sons Inc.

Inmon W.H., Imhoff Claudia y Sousa Ryan (1998) *Corporate Information Factory* EEUU: John Wiley & Sons Inc.

Inmon W.H. (1996) *Building the Data Warehouse (2ª Ed.)*. EEUU: John Wiley & Sons Inc.

Inmon, W.H. Strauss, D. Neushloss, G. (2008) DW 2.0: The Architecture for next generation of Data Warehousing. EEUU: Morgan Kaufman Series.

Krish Krishnan (2013) Data Warehousing in the Age of Big Data. The Morgan Kaufmann Series on Business Intelligence

Enlaces a internet

Página Openflights: <http://Openflights.org>

Página open data banco mundial: <https://data.worldbank.org/>

Búsqueda códigos IATA: <http://www.iata.org/publications/Pages/code-search.aspx>

Búsqueda por diferentes códigos (IATA, ICAO y FAA):
<http://www.avcodes.co.uk/>

SQL Server Analysis Services 2016 (SSAS) Tutorial

<http://www.mssqltips.com/sqlservertip/1167/getting-started-with-sql-server-analysis-services/>

Oracle Database 11g Express Edition Tutorial Series:

http://apex.oracle.com/pls/apex/f?p=44785:24:112757326496942::NO:24:P24_CONTENT_ID%2CP24_PREV_PAGE:5922%2C2

Getting started with Pentaho Data Integration:

<https://help.pentaho.com/Documentation/8.0/Setup/Evaluation/Tutorials/0C0>