

2016

# Bases de Dades Avançades

## PRÀCTICA 2

Magatzems de dades  
i Cubs per anàlisi  
multidimensional

Gestió de VolaUB

Karen Samsó  
Judit Domènech

Universitat de Barcelona



## Índex

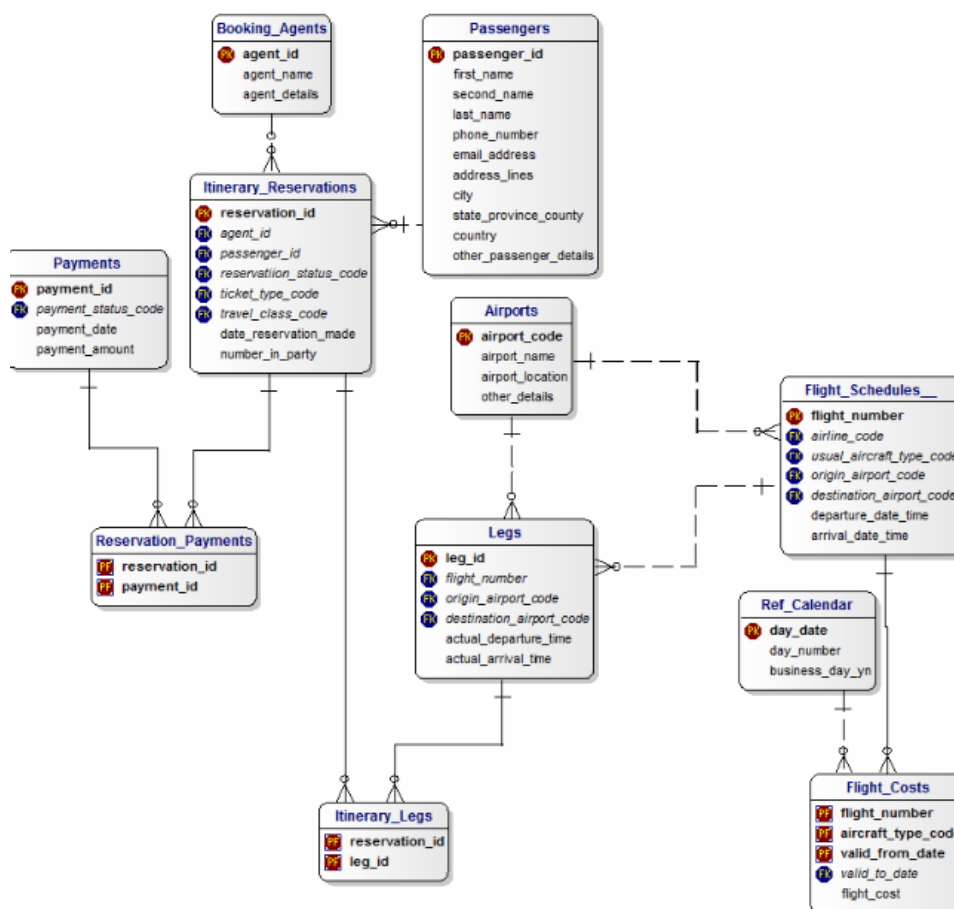
1. Introducció.....	3
2. DataMart .....	3
3. Cub .....	5
4. Problemes sorgits .....	5
Problemes en afegir nivells a la dimensió temporal .....	5
Problemes en voler fer una mesura a partir d'una propietat d'una dimensió .....	5

## 1. Introducció

En aquesta pràctica tenim com objectiu aprendre a dissenyar magatzems de dades (un DataMart) i a implementar cubs per a l'anàlisi multidimensional de les dades, a partir del motor OLAP Mondrian.

## 2. DataMart

En la primera part de la pràctica, es demana dissenyar el DataMart de VolaUB (que és el nom de la base de dades que implementarem en postgres) i afegir-hi algunes dades, a partir d'aquest model:

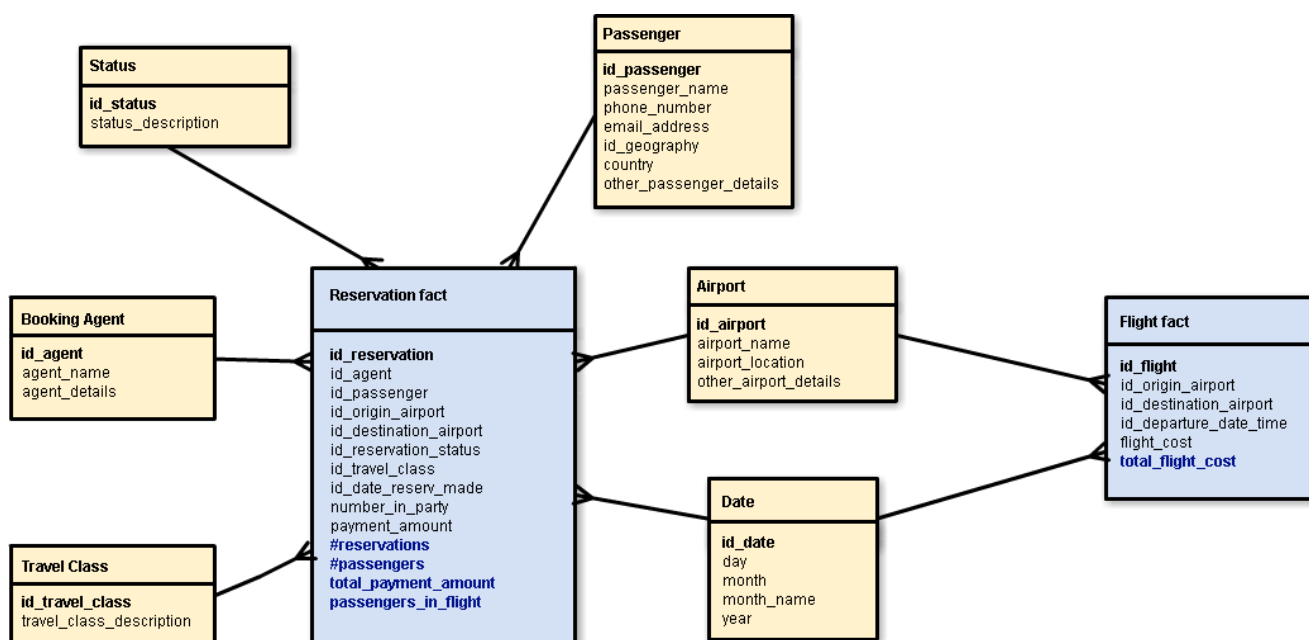


S'ha començat amb un disseny inicial senzill que s'ha anat modificant en visualitzar el cub i veure què es podia afegir o canviar per millorar-lo i ser capaces de mostrar més informació (o resoldre problemes que sorgien a l'hora de fer la implementació). D'aquesta manera el disseny ha anat variant al llarg de la pràctica de la següent manera:

- Es va començar dissenyant 3 taules de fets (Reservation, Payment i Flight), de manera que es podien diferenciar bé les parts que consideràvem més importants del model inicial.
  - Reservation contenia el passatger, l'agent, el status, la classe, la data de reserva i el número de persones per les que es fa;
  - Payment contenia el passatger, el status, la data de pagament i la quantitat;
  - Flight contenia l'aeroport d'origen, el de destí, la data de sortida del vol i el cost del vol.

- Es va reduir tot a una sola taula, eliminant Payment i Flight, deixant només Reservation de manera que aquesta contenia tota la informació que abans tenien les altres dues taules. Aquest canvi ens permetia fer consultes entre dimensions que no estaven directament relacionades en el disseny anterior (com per exemple, la relació entre un vol i un passatger), però alhora reduïa molt la possible informació que abans es tenia, fent un DataMart menys complet.
- Finalment, es va arribar a la conclusió que la taula Flight havia de ser una taula de fets a part i que Reservation podia contenir la informació relativa al pagament per simplificar.

El resultat final és aquest:



Consideracions importants:

- Reservation i Flight contenen una dimensió compartida (que representa tant l'aeroport d'origen com l'aeroport de destí), informació que es vincularà més tard per poder relacionar el cost del vol amb els pagaments, obtenint així un membre calculat que indiqui el benefici econòmic de la companyia.
- No s'ha inclòs la dimensió corresponent al model d'avió, ni la que representa una aerolínia, ja que s'ha considerat que el DataMart és per la gestió d'una sola companyia aèria i que no té cap rellevància l'avió concret que fa el vol per l'anàlisi de negoci que hem dissenyat.
- Les mètriques considerades en Reservation són: el nombre de reserves, el nombre de passatgers, el total de tots els pagaments realitzats i el total de tots els passatgers que aniran en un vol (el nombre de persones per les quals un passatger ha efectuat la reserva).
- S'ha considerat que l'única mètrica útil en Flight és el cost total de tots els vols.
- Passenger conté una geografia (representada per un identificador i un país) que permet fer un nivell més en la jerarquia, de manera que els passatgers es poden agrupar segons el país on viuen.

### **3. Cub**

En aquesta segona part, es demana dissenyar el cub per fer l'anàlisi multidimensional mitjançant l'Schema-Workbench de Mondrian.

Per fer la implementació, s'ha començat provant amb els diferents dissenys de DataMart comentats en l'apartat anterior, i mostrant només unes poques dimensions, de manera que s'ha anat ampliant la informació mostrada.

La dificultat va aparèixer a l'hora de voler fer membres calculats a partir de certes mètriques que no pertanyien al mateix cub, fet que ha comportat la majoria de canvis en el disseny, però s'ha acabat optant per afegir un cub virtual que relaciona els dos cubs (Reservation i Flight) per les dimensions que tenen en comú (Airport Origin i Airport Destination). Això permet obtenir una nova mètrica calculada que representa el benefici de la companyia restant el cost dels vols a tots els ingressos obtinguts corresponents ( $\text{Benefit} = \text{Total payment amount} - \text{Total flight cost}$ ).

En conclusió, el disseny del DataMart i els cubs implementats, ens permeten fer totes aquelles preguntes de negoci que impliquin les mètriques comentades per cadascuna de les dimensions que s'han considerat d'informació rellevant.

### **4. Problemes sorgits**

#### **Problemes en afegir nivells a la dimensió temporal**

La primera dimensió amb nivells que vam crear va ser la dimensió temporal. Vam afegir el nivell *Year* sense cap problema a l'hora de visualitzar el cub, però en afegir un segon nivell *Month*, es produïa un error quan intentàvem visualitzar la informació per mesos. La causa de l'error era que en les dimensions temporals, no es podia indicar el nivell mitjançant un string, s'havia de fer a partir d'un nombre enter. Un cop sabuda la causa de l'error, vam modificar l'script de creació de taules per tal que els mesos es guardessin de forma numèrica i d'aquesta forma ja no sorgia el problema.

#### **Problemes en voler fer una mesura a partir d'una propietat d'una dimensió**

Aquest és el problema que ens ha costat més solucionar. En el segon disseny que només contenia la taula de fets Reservation, teníem una dimensió Flight que tenia com a propietat *Flight Cost*, variable que volíem agafar per tal de calcular el benefici relacionant-la amb la mètrica *Total payment amount*. Quan s'intentava sortir errors del tipus:

- No poden haver una jerarquia i una dimensió utilitzant el mateix membre
- Mondrian Error:Named set in cube 'Reservation' has bad formula

És per això que es va optar a utilitzar un cub virtual amb el disseny definitiu comentat.

En aquesta entrega s'adjunta un Script que conté la creació de taules i una sèrie de inserts a la base de dades, l'esquema .XML de VolaUB i els diferents arxius .JSP que permeten visualitzar cada un dels cubs implementats.