

Bases de Dades Avançades

Practica 1 - Consultes Analítiques

Alumne
Professor

Germán Dempere Guillermo
Enric Biosca

Taula de Continguts

Objectius i Tecnologies	2
Desenvolupament	3
De Taules Relacionals a Taules de Fets	3
Diseny del Cub	5
Execució en l'Entorn Jetty	6
Conclusions	7

Objectius i Tecnologies

L'objectiu de la practica és poder realitzar consultes analitiques sobre un model de dades que ens permeti obtenir insights de les operacions econòmiques del nostre restaurant fictici.

Per arribar a aquest objectiu dissenyarem un esquema de bases de dades i un model de Cub OLAP que permeti aquestes consultes analitiques sobre un conjunt de dades. Aquest conjunt de dades existeixen en un altre model orientat a treballar amb aquestes dades de manera operacional, no analítica.

Farem servir com a suport físic de magatzem de dades permanent una base de dades MySQL en un servidor extern.

Per fer la traducció de queries MDX a consultes relacionals, dissenyarem un esquema amb el software Schema Workbench. Aquesta eina ens permet definir mesures i dimensions que aplicarem sobre un o més cubs OLAP. I es a través de consultes en l'espai del cub que podrem extreure les dades en el format que nosaltres volguem interpretar.

Per visualitzar la sortida de les consultes, iniciarem un servidor Jetty, que permet inserir els nostres esquemes i consultes en llenguatge XML i MDX. Jetty proporciona un entorn Mondrian que finalment executa les consultes.

Desenvolupament

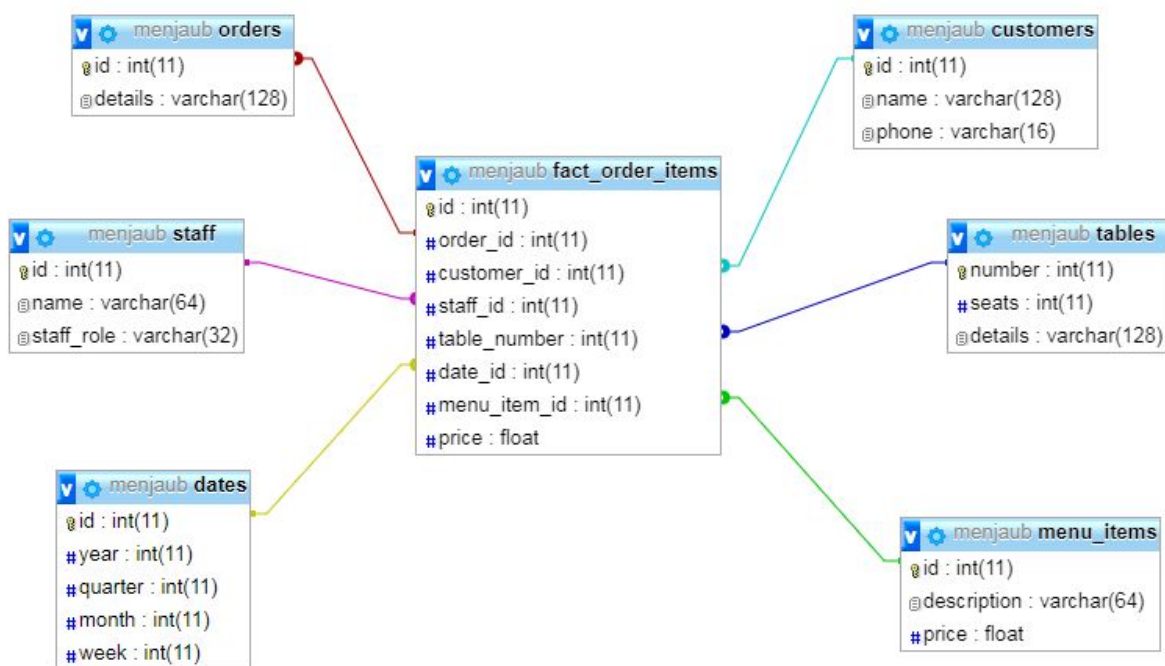
Encara que el desenvolupament estigui dividit en dos parts, els requisits i dependències d'una inevitablement influencien decisions de disseny a l'altra part. Degut a això, s'han fet varies iteracions del disseny a fi de perseguir una capacitat analítica i una flexibilitat majors. El producte final és el que es presenta a continuació.

De Taules Relacionals a Taules de Fets

El primer pas de la pràctica és interpretar i entendre el model de bases de dades que es proposa i extreure les unitats mínimes i les mesures que volem analitzar. En el nostre cas, es demanen dos requisits:

- Analitzar l'activitat econòmica del restaurant, les comandes, els empleats, i tots els demés aspectes del restaurat.
- Analitzar l'activitat de les reserves, quanta gent no es presenta, i les comissions dels diferents serveis de reserves.

Per cobrir la primera necessitat s'ha dissenyat el següent esquema de taula de *facts* en forma de Estrella:

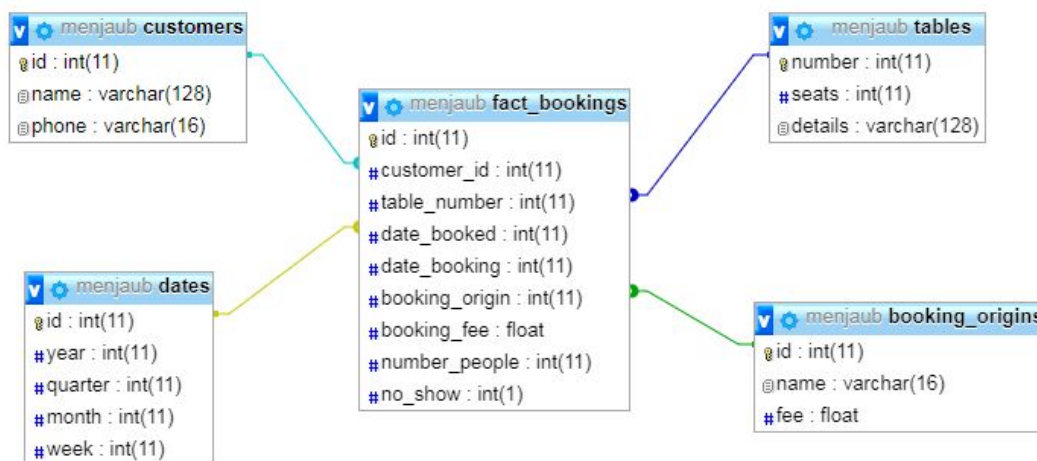


La decisió més important i complicada per poder escollir que considerem com a fet, es possiblement la **granularitat** que volem apreciar. En un principi es va considerar la granularitat a nivell de comanda, pero aquest model previ perdia molta informació que pot ser d'interés.

En una revisió posterior es va decidir interpretar un sol ítem de la comanda com a unitat de fet. De manera que no guardem comandes enteres a dins de la base de dades, sino cada cop que es ven una cosa. Seguim tenint una unitat de comanda, pero aquesta comanda pot correspondre a més d'un registre.

Aquesta elecció permet una flexibilitat major que solament comandes. També es va considerar la granularitat a nivell de ingredient. Pero com el preu es a nivell de ítem de menu i no de ingredient, es complicaven els models. També introduïa un nivell excésiu de fragmentació. Per consegüent considerem que a nivell d'ítem queda ampliament recollida la informació rellevant.

Per fer front al segon requisit, es dissenya el següent esquema:



En aquest cas, si que s'opta directament per agafar una reserva com a unitat mínima de granularitat, ja que no es pot subdividir en res més petit. Tal i com es demana als requisits, la taula de fets de la reserva incloïx el camp **no_show**, que en cas d'estar marcat com a 1, indica que la persona de la reserva no s'ha presentat.

Diseny del Cub

Paral·lelament al diseny de les taules, el diseny del cub OLAP es el que al final determinarà les capacitats analitiques del nostre sistema. Les dimensions i mesures del cub estan directament subordinades a com hem estructurat les taules en el pas anterior. Es per aquest fet que s'han redisenyat varies vegades les taules.

En la versió final podem trobar els dos cubs següents, amb les seves dimensions:

Order Items	Bookings
Dimensions Customer Order Table Staff Date Menu Item Measures Amount Orders Item Sales Customers	Dimensions Customer Table Booking Made (<i>Date</i>) Booking Date (<i>Date</i>) Origin Measures Fee People No Show Bookings No Show Ratio <i>CM</i>

I les dimensions tenen els següents nivells:

Date	Customer	Table	Staff
Year Quarter Month Week	Name	Seats Details	Role Name
Order	Menu Item	Bookings	
Details	Description	Origin	

Execució en l'Entorn Jetty

Despres de configurar correctament els fitxers d'esquema i JSPs, si accedim a <http://localhost:8080/mondrian>, tindrem enllaços directes a les queries analitiques per ser executades sobre el cub. Podem veure varis exemples a continuació:

Comandes del Q3 de 2016

	Measures			
default	Amount	Orders	Item Sales	Customers
-All Date.defaults	21,053.1	1,559	3,125	790
-2016	10,290.9	778	1,541	542
+1	2,503.9	190	365	175
+2	2,440.85	189	378	167
+3	2,576.55	193	379	169
+7	727.85	56	110	52
+8	836.45	68	132	65
+9	1,012.25	69	137	67
+4	2,769.6	206	419	185
+2017	10,762.2	781	1,584	549

Items que es demanen més a les comandes

	Measures			
default	Amount	Orders	Item Sales	Customers
-All Menu Item.defaults	21,053.1	1,559	3,125	790
Cake	1,197.7	365	406	304
Cesar Salad	1,732.5	364	385	304
Chicken drumsticks	2,567.5	367	395	315
Coca-cola	550.5	346	367	292
Hot coffe	866.25	351	385	299
Roasted Beef	5,115.25	366	395	303
Salad	1,576.05	358	399	298
Tenderloin	7,447.35	356	393	305

Origens de les reserves amb el ratio de gent que no es presenta

	Measures		
Origin.default	People	Fee	No show ratio
-All Booking Origin.defaults	1,264	1,498.8	0.16
affiliate	335	485.75	0.13
coupon	323	145.35	0.183
referral	292	569.4	0.17
website	314	298.3	0.159

Dates per les que es fan reserves

	Measures		
Booking Date.default	People	Fee	No show ratio
-All Date.defaults	1,264	1,498.8	0.16
-2016	611	744.45	0.182
+1	81	96.95	0.212
+2	181	218.95	0.157
+3	161	209.45	0.162
+4	188	219.1	0.211
+10	53	58.35	0.143
+11	66	80.7	0.261
+12	69	80.05	0.222
+2017	653	754.35	0.14

Conclusions

Podem dir que hem arribat a cobrir els objectius que ens havíem plantejat. A través d'una sèrie de consultes analítiques podem obtenir insights en com està funcionant el nostre restaurant.

Això ens permet identificar per exemple, de les reserves, quins orígens no estan funcionant bé, o causen uns costos alts. O per part de les comandes, quins ítems són més populars, i quins, encara que no siguin tan populars (menys nombre de clients), donen potser més diners que aquells que són més demanats però no són tan cars.

Totes aquestes preguntes serien més complicades de realitzar si fèssim servir el model de base de dades operatiu. És a dir, el model que permet gestionar el negoci amb tots els detalls i amb el seu dia a dia de comandes.

Es per això que s'estableix una diferència molt clara entre bases de dades orientades a guardar informació operativa, i bases de dades analítiques, que prenen una forma des-normalitzada a propòsit, trencant tot el que a priori s'ensenya a BBDD1, en favor de la funcionalitat molt més orientada a ser consultades i analitzades.