



**Universidad Católica de Santiago del Estero**

Departamento Académico Rafaela

Ingeniería en Informática

---

# Universidad Católica de Santiago del Estero

**Carrera:** Ingeniería en Informática.

**Asignatura:** Business Intelligence & Big Data.

**Docentes:**

- Aguirre, Alejandro.
- D'lorio, Lorena.
- Vera, Marcela.

**Tema:** Data Warehouse y Data Mining.

**Alumnos:**

- Barsotti, Alejandro.
- Pellegrinet, Nicolás.
- Sclerandi, Wendy.

**Ciudad:** Rafaela.

**Fecha de entrega:** 23/10/2017.



## **Índice**

Tiempo utilizado en la resolución del trabajo práctico .....	3
Dificultades .....	3
Supuestos.....	4
Esquemas Dimensionales .....	5
Reportes solicitados .....	5
Modificaciones necesarias ante nuevos requerimientos .....	12

## **Tiempo utilizado en la resolución del trabajo práctico**

Se presenta una tabla con el detalle de las actividades realizadas como parte de la resolución del trabajo práctico propuesto y la cantidad de horas que fueron utilizadas para tal fin.

Actividad	Cantidad de Horas
Armado de tablas para bases de datos operacionales	4
Búsqueda de datos de productos	5
Búsqueda de datos de actividades	1
Procesamiento y generación de datos	9
Creación de bases de datos operacionales	4
Investigación y prueba de backup/restore de bases de datos	6
Ordenado de scripts de inserción de datos	1
Creación de Dimensiones y Hechos	3
Consultas y revisión general	6
Investigación y generación de transformaciones en Spoon	18
Investigación y creación de cubos OLAP	8
Investigación, prueba y generación de reportes en Pentaho	17
Generación de Informe	6
Revisión general	2
<b>Total</b>	<b>3 días y 14 horas</b>

Es importante mencionar que, si bien el tiempo efectivo fue de 3 días y 14 horas, debido a diversas obligaciones académicas y algunas obligaciones extras en cuanto a la universidad se refiere, la realización del trabajo práctico se llevó a cabo de forma distribuida alrededor de 12 días, aproximadamente.

## **Dificultades**

A continuación, se procede a listar las dificultades presentadas en la realización del presente trabajo práctico:

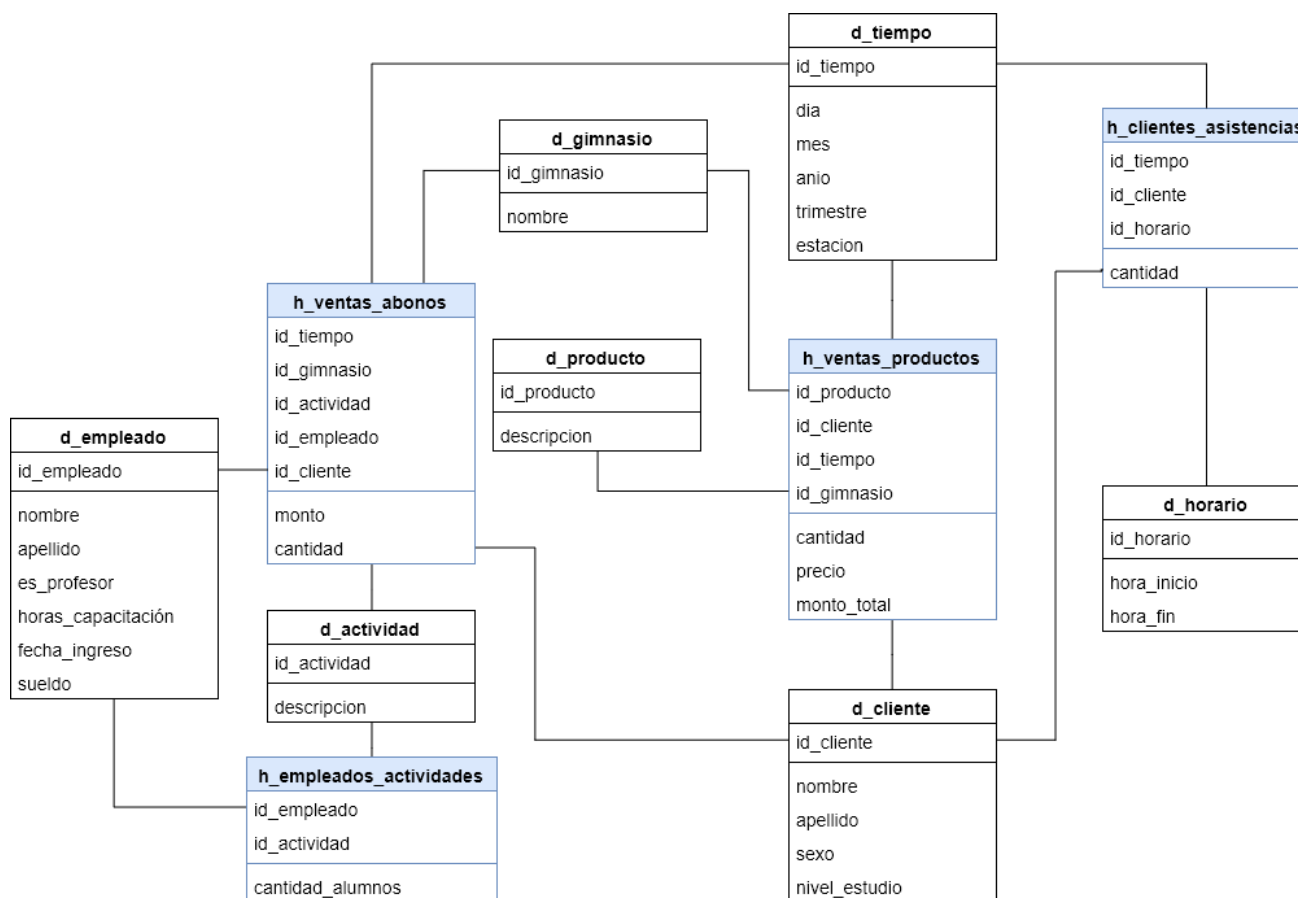
- Falta de conocimiento y soporte online en el uso de las herramientas proporcionadas.
- Complicaciones para generar backups de cada uno de los puntos planteados. Si bien el proceso es intuitivo y se encuentran tutoriales en internet, tuvimos demasiadas dificultades para lograr que el proceso de restauración funcione correctamente. Esta actividad significó un gasto extra en cuanto a horas dedicadas al trabajo se refiere.
- Debido a la necesidad de generar datos para las bases de datos operacionales, gran parte del tiempo total dedicado a la resolución del trabajo práctico fue destinado a la generación de scripts de carga de datos para las mismas (aproximadamente 4 días). Esta tarea incluyó: búsqueda de volúmenes de datos en internet, desarrollo y ejecución de los scripts generadores, generación y ejecución de scripts de inserción, corrección de inconsistencias en los datos.

- Dificultad en la comprensión de algunos requerimientos solicitados en el enunciado del trabajo práctico, puesto que el mismo contaba con algunos supuestos que a simple vista no eran posibles de inferir.
- Inconvenientes a la hora de tratar de trabajar paralelamente, principalmente debido a que todo debe realizarse en una misma máquina virtual. Si bien cada integrante del grupo efectuaba el armado de las bases de datos en su propia máquina virtual, ciertas tareas no podían paralelizarse. Esto generó que el ritmo de trabajo no sea el ideal.
- Dificultad en el diseño de tablas de dimensiones y hechos para el Data Warehouse. Contar con más práctica sobre ejercicios de este tipo en la cátedra hubiera sido de ayuda para la resolución de los conflictos.
- Necesidad de investigar el funcionamiento de las herramientas de Pentaho.
- Dificultad en la creación de los cubos, debido a la falta de conocimiento de este tema, y a diversos errores que surgieron en la máquina virtual durante la realización de los mismos.
- Escaso tiempo para la resolución del trabajo.

## **Supuestos**

- La ubicación geográfica de los empleados se determina por el campo *gimnasio* en la tabla *Empleado* del *Sistema de Recursos Humanos*, ya que los mismos solo pueden trabajar en un único gimnasio.
- En el código de las ventas (de la forma “GIM[nro gim]-[código ticket]”) se indica a qué gimnasio pertenece cada una. Esto aplica al *Sistema de Abonos* y a los sistemas de *Facturación de Productos*.
- La ubicación geográfica de los clientes es determinada por el código de gimnasio que tienen los tickets de las compras que realizaron, según lo indicado en el supuesto anterior.
- Se consideran ofertas de actividades aquellas cuyo código de actividad sea impar.
- Se consideran productos de oferta aquellos cuyo código de producto sea impar.
- Se agrega la tabla *Nivel\_estudio* en el *Sistema de Abonos*, con código y descripción.
- Se agrega el campo *hora* a la tabla *Asistencia* del *Sistema de Abonos* para determinar a qué actividad asiste el cliente en la fecha indicada y, asimismo, poder detectar los horarios de mayor concurrencia al gimnasio.
- Las actividades en un gimnasio determinado son dictadas exclusivamente por un instructor dado. Asimismo, un instructor no puede dictar más de una actividad.
- Todos los gimnasios dictan todas las actividades definidas en la tabla de *Actividades* del *Sistema de Recursos Humanos*.

## Esquemas Dimensionales



## Reportes solicitados

A continuación, se detallan los resultados de los reportes de información solicitados por el nivel gerencial y los analistas de negocio de la empresa.

*Nota: las tablas aquí exhibidas pueden no representar el resultado del reporte completo, sino una sección representativa de los mismos a modo de ejemplo. Los reportes originales generados pueden ser consultados en la carpeta "Reportes" adjunta en la entrega del presente trabajo práctico.*

### Cantidad de ventas de abonos por año y mes

Año: 2016 - Mes: 1	
Actividad	Cantidad de ventas
Body pump	169
Circuito switching	176
Clases de body balance	181

Clases de body combat	221
Clases de ciclo indoor	188
Clases de cross training - intensity	160
Clases de foam roll - estiramientos	181
Clases de gap (glúteos, abdominales y piernas)	181
Clases de hiit - entrenamiento de alta intensidad	167
Clases de kick boxing	185
Clases de multifitness	203
Clases de pilates	179
Clases de strong by zumba	194
Clases de zumba	205
Ejercicio cardiovascular y fuerza	191
Entrenamiento del core	179
Glúteos up	214
On funcional	181
Tbc - total body conditioning	186
<b>Total</b>	<b>3541</b>

Cantidad de ventas de abonos por año y estación

<b>Año: 2016 - Estación: invierno</b>	
<b>Actividad</b>	<b>Cantidad de Ventas</b>
Body pump	531
Circuito switching	516
Clases de body balance	555
Clases de body combat	549
Clases de ciclo indoor	560
Clases de cross training - intensity	550
Clases de foam roll - estiramientos	532
Clases de gap (glúteos, abdominales y piernas)	542
Clases de hiit - entrenamiento de alta intensidad	578
Clases de kick boxing	532
Clases de multifitness	559
Clases de pilates	564
Clases de strong by zumba	552
Clases de zumba	536
Ejercicio cardiovascular y fuerza	569
Entrenamiento del core	536
Glúteos up	538

On funcional	542
Tbc - total body conditioning	569
<b>Total invierno</b>	<b>10410</b>

Cantidad de ventas de abonos por año y trimestre

<b>Año: 2016 - Trimestre: 1° Trimestre</b>	
<b>Actividad</b>	<b>Cantidad de Ventas</b>
Tbc - total body conditioning	572
Clases de strong by zumba	606
Glúteos up	586
Clases de body combat	588
Clases de foam roll - estiramientos	553
Clases de pilates	569
Clases de body balance	595
Clases de ciclo indoor	583
Clases de kick boxing	556
Clases de multifitness	585
Body pump	576
Ejercicio cardiovascular y fuerza	583
Clases de zumba	563
Entrenamiento del core	555
Clases de hiit - entrenamiento de alta intensidad	521
Clases de cross training - intensity	523
Clases de gap (glúteos, abdominales y piernas)	561
On funcional	528
Circuito switching	546
<b>Total 1° Trimestre</b>	<b>10749</b>

Cantidad de ventas de abonos por gimnasio

<b>Gimnasio: GIM 1</b>	
<b>Actividad</b>	<b>Cantidad de Ventas</b>
Body pump	495
Circuito switching	441
Clases de body balance	517
Clases de body combat	524
Clases de ciclo indoor	522
Clases de cross training - intensity	493
Clases de foam roll - estiramientos	475

Clases de gap (glúteos, abdominales y piernas)	481
Clases de hiit - entrenamiento de alta intensidad	533
Clases de kick boxing	498
Clases de multifitness	555
Clases de pilates	485
Clases de strong by zumba	518
Clases de zumba	493
Ejercicio cardiovascular y fuerza	513
Entrenamiento del core	501
Glúteos up	487
On funcional	511
Tbc - total body conditioning	519
<b>Total GIM 1</b>	<b>9561</b>

Cantidad de ventas de abonos por instructor

Empleado	Cantidad de Ventas
Maximiliano Moriondo	358
Gaston Emanuel Bodoira	364
Claudia Viviana Fleitas	364
Maria Alejandra Barrera	369
Ariel Gustavo Pascutti	370
Maria Josefina Moyano	371
Daniel Julio Taborda	372
Mauricio Angel Cornier	376
Alfredo Sebastian Achkar	379
Horacio German Favre	380
Federico Giancarelli	381
Marco Esteban Sartor	381
Matias Alejandro Fernandez	383
Juan Marcelo Muchiut	386

Cantidad de ventas de abonos por actividad física

Actividad	Cantidad de Ventas
Body pump	4379
Circuito switching	4344
Clases de body balance	4449
Clases de body combat	4401
Clases de ciclo indoor	4487



Clases de cross training - intensity	4427
Clases de foam roll - estiramientos	4328
Clases de gap (glúteos, abdominales y piernas)	4364
Clases de hiit - entrenamiento de alta intensidad	4485
Clases de kick boxing	4389
Clases de multifitness	4436
Clases de pilates	4466
Clases de strong by zumba	4367
Clases de zumba	4389
Ejercicio cardiovascular y fuerza	4413
Entrenamiento del core	4479
Glúteos up	4390
On funcional	4335
Tbc - total body conditioning	4403

Horarios de mayor concurrencia a los gimnasios

Horario	Cantidad de Asistencias
10:00 hs a 11:00 hs	96503
11:00 hs a 12:00 hs	78251
12:00 hs a 13:00 hs	50701
13:00 hs a 14:00 hs	94443
14:00 hs a 15:00 hs	67471
15:00 hs a 16:00 hs	79204
16:00 hs a 17:00 hs	164759
17:00 hs a 18:00 hs	68422
18:00 hs a 19:00 hs	93668
19:00 hs a 20:00 hs	110626
20:00 hs a 21:00 hs	44716
21:00 hs a 22:00 hs	98733
8:00 hs a 9:00 hs	108548
9:00 hs a 10:00 hs	80203
<b>Total</b>	<b>1236248</b>

Cantidad de asistencias por cliente

Cliente	Cantidad de Asistencias
Alicia Estela Holzer	42
Pablo Armando Mendez	94
Nancy Leticia Sosnowski	84



Ricardo Vicente Peralta	48
Nicolas Alejandro Badino	55
Daniel Horacio Gallo	46
Franco Gustavo Rufino	31
Marcela Lilian Romero	11
Juan Pablo Colautti	7
Romina Vanesa Ferrando	90
Lucio Adrián Carnevalle	20
Cristian Pablo Gulli	54
Juan Alejandro Rosso	39
Agustín Andrés Gonzalez Moya	130

Cantidad de alumnos por instructor y horas de capacitación

Empleado	Actividad	Horas de Capacitación Totales	Cantidad de Alumnos
Facundo Miguel Angel Portillo	Entrenamiento del core	238	463
Juan Jose Suhit	Tbc - total body conditioning	262	431
Franco Antonio Ramón Brest	Clases de hiit - entrenamiento de alta intensidad	152	441
Diego Luis Engler	Clases de gap (glúteos, abdominales y piernas)	210	445
Victor Hugo Castillo	Clases de zumba	216	493
Carla Lorena Salamone	Clases de ciclo indoor	25	474
Marcelo Ramon Costanzo	Clases de gap (glúteos, abdominales y piernas)	199	416
Marcelo Alejandro Hosch	Clases de foam roll - estiramientos	92	455
Aldo Iñiguez	Clases de foam roll - estiramientos	44	419
Juan Ignacio Vanney	Clases de kick boxing	126	387
Federico Giancarelli	Clases de cross training - intensity	227	381
Gonzalo Alfredo Roman	Clases de body combat	293	457

Diego Gabriel Giraudo	Clases de ciclo indoor	143	495
Edgardo Jose Nazzo	Circuito switching	176	480

Cantidad de ventas de productos por gimnasio

<b>Gimnasio: GIM 1</b>		
<b>Producto</b>	<b>Cantidad Vendida</b>	<b>Monto Total</b>
Correa de agarre	62	\$ 11.370
Cinturón super powerlifter	68	\$ 34.694
Crema reductora - quema grasa	82	\$ 53.245
Lady pro active gloves	71	\$ 30.405
Energyplus	43	\$ 26.829
Carnitin diet bar	98	\$ 102.811
Maillot - culotte corto bike multipower	86	\$ 176.038
Tostadas de pan integral	91	\$ 1.092
Shaker hsn sports powerexplosive blanco	77	\$ 9.293
Transporter duffel	84	\$ 125.472
Lady gel gloves	80	\$ 35.898
Metactif	77	\$ 62.509

Ingresos de abonos según sexo y nivel de estudio

<b>Sexo: Femenino</b>	
<b>Nivel de Estudio</b>	<b>Monto total</b>
Primario completo	\$ 1.382.568.420
Primario incompleto	\$ 1.459.831.430
Secundario completo	\$ 1.398.956.670
Secundario incompleto	\$ 1.407.196.014
Terciario completo	\$ 1.311.785.688
Terciario incompleto	\$ 1.524.379.896
Universitario completo	\$ 1.516.324.016
Universitario incompleto	\$ 1.662.173.311
<b>Total Femenino</b>	<b>\$ 11.663.215.445</b>

Cantidad de ventas de abonos en oferta

<b>Tipo de Precio: Oferta</b>	
<b>Actividad</b>	<b>Cantidad Vendida</b>
Clases de kick boxing	4389

Body pump	4379
Clases de hiit - entrenamiento de alta intensidad	4485
On funcional	4335
Entrenamiento del core	4479
Clases de body combat	4401
Clases de gap (glúteos, abdominales y piernas)	4364
Glúteos up	4390
Clases de zumba	4389
Clases de strong by zumba	4367
<b>Total Oferta</b>	<b>43978</b>

#### Cantidad de ventas de productos en oferta

Tipo de Precio: Oferta	
Actividad	Cantidad Vendida
Botines adidas messi 16.1 fg	245
Botines adidas nemeziz 17.1 fg	234
Botines nike mercurialx victory vi cr7 tf	226
Carboprotein ultra	226
Carboprotein ultra	258
Carnitín plus	249
Cinturón extra support	254
Cinturón spibelt energy	234
Competitor backpack	241
Correas entrenamiento scitec	239
Creaplus - creatina monohidrato	221
Creaplus - creatina monohidrato	226
Crema reductora - quema grasa	224

### **Modificaciones necesarias ante nuevos requerimientos**

Nuevo requerimiento: “*Si Zumba Gym decidiera incorporar información acerca del stocks disponible de sus productos en el data warehouse. El stock deberá estar disponible a nivel de producto, día y local.*”

Modificaciones necesarias:

- Agregar una tabla “Stock” a las bases de datos operacionales de facturación de productos (Facturación 1 y Facturación 2), que indique la cantidad en stock de un producto dado al finalizar el día en el gimnasio correspondiente al sistema.

- Agregar al Data Warehouse una nueva tabla de hechos (“H\_Stock”) con los siguientes campos: id\_producto, id\_tiempo, id\_gimnasio, cantidad\_stock.

Nuevo requerimiento: *“Si se lanzara una tarjeta de fidelización de clientes y se pudiera identificar a qué cliente se hizo cada venta (tanto de abonos como de otros tipos de productos).”*

Modificaciones necesarias:

- Los clientes actualmente ya son identificados en cada venta, por lo que no es necesario efectuar ningún cambio al respecto.
- Si se quisiera llevar un registro de los clientes que tienen o no tarjeta, y su respectivo número, debería agregarse el campo “tarjeta\_fidelización” a las tablas de clientes correspondientes.

Nuevo requerimiento: *“Poder evaluar el impacto que tiene en las ventas, la antigüedad, capacitación y el salario de los empleados.”*

Modificaciones necesarias:

- De no haber sido incluidos los datos “salario” y “fecha de ingreso” en la dimensión “Empleado” del Data Warehouse, estos deberían ser agregados.
- Generar dos nuevos reportes (uno para ventas de abonos y otro para ventas de productos) que presenten la cantidad de productos o abonos vendidos por empleado, incluyendo información sobre la antigüedad, salario y capacitación del vendedor.

Nuevo requerimiento: *“De cada local es necesario poder analizar las ventas (importe y cantidad) con respecto a la cantidad de actividades definidas y los metros cuadrados que disponen.”*

Modificaciones necesarias:

- Agregar tabla general centralizada con información de todos los locales de la empresa, donde se incluya el dato “metros cuadrados”.
- Agregar dato “metros cuadrados” a la dimensión “d\_Gimnasio” del Data Warehouse para poder visualizar esta información en el reporte de ventas.