



# NOVÁTICA

UPGRADE

Revista de la Asociación  
de Técnicos de Informática

Nº 167, enero-febrero 2004, año XXX

Redes inalámbricas:  
una nueva era en las telecomunicaciones

UPGRADE  
Novática, revista fundada en 1975 y decana de la prensa informática española, es el órgano oficial de expresión y formación continua de ATI (Asociación de Técnicos de Informática). Novática edita también Upgrade, revista digital de CEPIS (Council of European Professional Informatics Societies), en lengua inglesa.

<http://www.ati.es/novatica/>  
<http://www.upgrade-cepis.org/>

ATI es miembro fundador de CEPIS (Council of European Professional Informatics Societies) y tiene un acuerdo de colaboración con ACM (Association for Computing Machinery). Tiene asimismo acuerdos de vinculación o colaboración con AdSpain, AIZ y ASTIC.

### CONSEJO EDITORIAL

Antoni Carbo Nal / Miquel Ferrer, Francisco López Crespo, Julián Marcelo Cocho, Celestino Martín Alonso, José Mojás i Bertrán, Roberto Moya Oulles, César Pérez Chirinos, Mario Piattini Velhuis, Fernando Plaza Gómez (Presidente del Consejo), Miquel Ferrer, Asociación Murze Herranz

### Coordinación Editorial

Rafael Fernández Calvo <rcalvo@ati.es>  
Composición y redacción  
Jorge Ullcer  
Traducciones  
Grupo de Lengua e Informática de ATI <http://www.ati.es/gf/lengua-informatica/>  
Administración  
Tomas Bunete, María José Fernández, Enric Camarero, Felicidad López

### SECCIONES TÉCNICAS: COORDINADORES

**Administración Pública Electrónica**  
Gonzalo García Arribas, Francisco López Crespo (MAP) <garcia.gonzalo@map.es>, <flore@ati.es>  
**Arquitecturas**  
Jordi Tórrilla (DAC-UPC) <jtorri@ac.upc.es>  
Victor Vinas Yébenes (Univ. de Zaragoza) <vvinas@unizar.es>  
**Avanzada BITIC**  
Marina Touriño, Manuel Palao (ASIA) <manuel@palao.com>  
**Bases de Datos**  
Coral Calero Muñoz, Mario G. Piattini Velhuis (Escuela Superior de Informática, UCLM) <coralcalero@uclm.es>, <mpiatini@int-cr.uclm.es>  
**Búsqueda y tecnologías**  
Isabel Hernández Collazos (Fac. Dechea de Donostia, UPV) <ihernando@legalek.net>  
Isabel Daniela Fernández de Marcos (Baviera & Davara) <idavara@davara.com>  
**Caracterización de la Informática**  
Joaquín Espinosa Moreno (CPS-UZAR) <espejeta@posta.unizar.es>  
Cristóbal Pareja Flores (DSIP-UCM) <pareja@dsip.ucm.es>  
**Estadística y Estadística**  
Jen-Rue Song (Korea, Ernst & Young) <jsong@ecc.edu>  
**Informática y Filosofía**  
José Carlos Uzcátegui (UPV) <jcarroz@unica.edu>  
Esperanza Marco (ESCET-URJC) <esmarco@escet.urjc.es>  
**Informática Gráfica**  
Miguel Chover Sellés (Universidad Jaume I de Castellón) <mchover@uji.es>  
Roberto Viro (Ergographes, sección española) <viro@dsic.upv.es>  
**Ingeniería del Software**  
Javier Dalaco Casin (ULSI-UPV) <dalaco@ulsi.ehu.es>  
Luis Fernández (IPHS-E-UEM) <lfern@iphs.esi.uem.es>  
**Inteligencia Artificial**  
Federico Barber, Vicente Boti (DSIC-UPV) <vboti@ulsi.ehu.es>  
**Interacción Persona-Computador**  
Juli Abascado González (PLA-UPV) <julio@si.ehu.es>  
Jesus Lopez Vidia (Univ. de Lleida) <jlovid@sup.udl.es>  
**Internet**  
Alfonso Alvarez Garcia (UD) <alvarez@ati.es>  
Llorenç Pages Casas (India) <pages@ati.es>  
**Lenguaje e Informática**  
M. del Carmen Ugarte (UDM) <cugarte@ati.es>  
**Lenguajes Informáticos**  
Andrés María López (Univ. Carlos III) <amarin@i.uclm.es>  
J. Ángel Velázquez (ESCET-URJC) <avelazqui@escet.urjc.es>  
**Literales e Informática**  
Alfonso Espaloso (FIR-Univ. de La Laguna) <alescolan@ull.es>  
**Lingüística computacional**  
Xavier Gomez Guzmán (Univ. de Vigo) <xgg@uvigo.es>  
María del Palomar (Univ. de Alicante) <mpalomar@dis.ua.es>  
**Mundo estudiantil**  
Adolfo Vázquez Rodríguez (Rama de Estudiantes del IEEE-UCM) <advazquez@ieee.org>  
**Profesión Informática**  
Rafael Fernández Calvo (ATI) <rcalvo@ati.es>  
Miquel Ferrer (Agrup. de Barcelona) <mferrer@ati.es>  
**Redes y servicios telemáticos**  
Luis Guisasa Coloma (DCOM-UPV) <lguisasa@dccm.upv.es>  
Josep Sala Pareta (DAC-UPC) <pareta@ac.upc.es>  
**Seguridad**  
Javier Arletio (Redes y Sistemas, Bilbao) <jarletio@orion.eusko.es>  
Javier López Muñoz (DSI Informática-UMA) <jlm@cc.uma.es>  
**Sistemas de Tiempo Real**  
Alejandro Alonso Juan Antonio de la Puente (DI-UPM) <jaalonso@puente@di.upm.es>  
**Software Libre**  
Jesús M. González (Barcelona, Padri de las Heras Quirós (SYSC-URJC) <jmgonzalez@sysc.escet.urjc.es>  
**Tecnología de Objetos**  
Jesus Garcia Mora (DIS-UM) <jgmora@correo.um.es>  
Gustavo Rossi (UR-UNESP, Argentina) <grosi@poli.info.unpl.edu.ar>  
**Tecnologías para la Educación**  
Juan Manuel Dodero Beardo (UCM) <dodero@inf.ucm.es>  
Francisco Riera (EMCA) <friera@warado.es>  
**Tecnologías y Lenguaje**  
Pablo Hernández Mesriano (Bluewin) <pablohm@bluewin.ch>  
**TIC para la Sanidad**  
Beltrán Masero Vargas (DI-UNEX) <vmasero@unex.es>  
**TIC y Turismo**  
Andrés Aguayo Maldonado, Antonio Guevara Plaza (Univ. de Mataga) <agayov, guevara@tic.uma.es>

Las opiniones expresadas por los autores son responsabilidad exclusiva de los mismos. Novática permite la reproducción de todos los artículos, salvo los marcados con © o copyright, debiéndose en todo caso citar su procedencia y enviar a Novática un ejemplar de la publicación.

**Coordinación Editorial, Redacción Central y Redacción ATI Madrid**  
Pacilla 65, 3º dcha, 28006 Madrid  
Tfn: 914029391; fax: 913093655 <novatica@ati.es>  
**Composición, Edición y Redacción ATI Valencia**  
Av. del Peño de Valencia 23, 46005 Valencia  
Tfn: fax: 963330392 <secretari@ati.es>  
**Administración y Redacción ATI Cataluña**  
Via Llobregat 41, P: 1º, 08003 Barcelona  
Tfn: 334125235; fax: 934127713 <secretari@ati.es>  
**Redacción ATI Andalucía**  
Rosal, Newton, s/n, Ed. Sadet,  
Isla Cartuja 41092 Sevilla, Tfn: fax: 954460779 <secretari@ati.es>  
**Redacción ATI Aragón**  
Lagasca 9, 3-B, 50005 Zaragoza  
Tfn: fax: 976235131 <secretari@ati.es>  
**Redacción ATI Asturias-Cantabria** <ga-asturcant@ati.es>  
**Redacción ATI Castilla-La Mancha** <ga-clmancha@ati.es>  
**Redacción ATI Galicia**  
Recinto Ferrial s/n, 36540 Silleda (Pontevedra)  
Tfn: 986551413; fax: 986550102 <secretari@ati.es>  
**Subscripción y Ventas**  
<http://www.ati.es/novatica/interes.html>, o en ATI Cataluña o ATI Madrid  
**Publicidad**  
Pacilla 65, 3º dcha, 28006 Madrid  
Tfn: 914029391; fax: 913093655 <novatica.publicidad@ati.es>  
**Imprenta**  
© Impresión S.A. Juan de Austria EE, D6005 Barcelona  
Modulito la pal. B 15, 15-19/75 - ISSN: 0211-2124; CODEN NOVAFEC  
Portada: Antonio Crespo Fox, © ATI 2004  
Diseño: Fernando Agresta / © ATI 2004

### editorial

#### ATI contra el cánón privado sobre soportes digitales

en resumen

#### Redes e historia

Rafael Fernández Calvo

### monografía

#### Redes inalámbricas: una nueva era en las Telecomunicaciones (En colaboración con Upgrade).

Editores invitados: M. Ufuk Çağlayan, Vicente Casares Giner y Jordi Domingo i Pascual

#### Presentación. Redes de acceso inalámbricas:

##### hacia las comunicaciones móviles integradas > 03

Vicente Casares Giner, Jordi Domingo Pascual

##### ¿Cuál es la extensión óptima de un enlace inalámbrico? > 06

M. Ufuk Çağlayan, Fikret Sivrikaya, Bülent Yener

##### Capacidad en Sistemas Celulares WCDMA: Métodos de Análisis > 11

Luis Mendo Tomás

##### Perspectiva de la gestión de recursos radio en las redes celulares > 15

Oriol Sallent Roig, Jordi Pérez Romero, Ramón Agustí Comes

##### Estrategias de Gestión de Localización en la próxima generación de Sistemas de Comunicaciones Móviles > 20

Pablo García Escalle, Vicente Casares Giner

##### Movilidad IP: macromovilidad, micromovilidad, calidad de servicio y seguridad > 28

Josep Mangues Bafalluy, Albert Cabellos Aparicio, René Serral Gracià, Jordi Domingo Pascual, Antonio Gómez Skarmeta, Tomás P. de Miguel, Marcelo Bagnulo, Alberto García Martínez

##### Redes Inalámbricas ad hoc como tecnología de soporte para la Computación Ubicua > 33

Juan Carlos Cano Escrivá, Carlos Miguel Tavares Calafate, Manuel José Pérez Malumbres, Pietro Marzoni

##### Las WPAN en el trayecto hacia la 4G > 39

Ramón Agüero Calvo, Johnny Choque Ollachica, José Ángel Irastorza Teja, Luis Muñoz Gutiérrez, Luis Sánchez González

### secciones técnicas

#### Bases de Datos

##### Diseño lógico de Almacenes de Datos: efectividad del diseño en estrella > 44

Coral Calero Muñoz, Mario Piattini Velhuis, Manuel A. Serrano Martín

#### Enseñanza Universitaria de la Informática

##### La Asociación de Enseñantes Universitarios de Informática (AENU) > 47

Pedro Blesa Pons, Joe Miró Julià, Francisco Ruiz González

#### Informática Gráfica

##### Visualización de terreno en tiempo real > 50

Cristina Rebollo Santamaría, Inmaculada Remolar Quintana, Miguel Chover Sellés

#### Ingeniería del Software

##### Buscando el Santo Grial de la Ingeniería del Software > 54

Robert L. Glass

#### Interacción Persona-Computador

##### Patrones de interfaz de usuario para la navegación orientada a objetos > 55

Pedro Juan Molina Moreno, Ismael Torres Boigues, Oscar Pastor López

#### Redes y servicios telemáticos

##### Análisis y Diseño de Políticas de Control de Admisión en Redes Celulares Multiservicio > 61

Vicent Pla Bosca, Vicente Casares Giner

#### Seguridad

##### Referencias autorizadas > 68

### sociedad de la información

#### programar es crear

##### Subcadenas en la secuencia "mira-y-di" (CUPCAM 2003, problema C, enunciado) > 73

Óscar Martín Sánchez, Manuel Carro Liñares

##### Reconstrucción de árboles inclinados a partir de dos de sus recorridos (CUPCAM 2003, problema B, solución) > 74

Cristóbal Pareja Flores, Ángel Herranz Nieva

### asuntos internos

#### Coordinación editorial / Programación de Novática > 76

#### Normas de publicación para autores / Socios Institucionales > 77

Coral Calero Muñoz, Mario Piattini Velthuis, Manuel A. Serrano Martín  
 Grupo ALARCOS, Universidad de Castilla-La Mancha

<{Coral.Calero, Mario.Piattini, Manuel.Serrano}@uclm.es>

# Diseño lógico de Almacenes de Datos: efectividad del diseño en estrella

Esta investigación es parte del proyecto CALDEA (TIC 2000-0024-P4-02) financiado por la Subdirección General de Proyectos de Investigación, Ministerio de Ciencia y Tecnología.

## 1. Introducción

Los almacenes de datos (*Data Warehouses*) surgieron por la necesidad de las organizaciones de contar con mecanismos que ayudaran a la toma de decisiones. Los almacenes de datos se han convertido en la tendencia más importante de la informática empresarial ya que proporcionan información muy relevante y precisa para mejorar las decisiones estratégicas, contando también con una alta inversión económica, que podría llegar a alcanzar en los próximos años, los 12 millones de dólares americanos [1].

Muchas organizaciones se sorprenden al comprobar que una que el almacén de datos ha sido diseñado correctamente es muy flexible y reutilizable [2]. Sin embargo, es fundamental asegurar que el diseño se ha realizado de forma correcta.

El modelado dimensional es la técnica normalmente utilizada para el modelado lógico de los almacenes de datos, en lugar de la utilización del modelo E/R y su transformación al modelo relacional. El modelado dimensional trata de presentar los datos en un marco estándar que sea intuitivo y con alto rendimiento en los accesos. Cada modelo dimensional está compuesto por una tabla de hechos y por varias tablas de dimensiones. La tabla de hechos es la tabla principal y contiene por un lado las medidas del negocio y por el otro un conjunto de claves ajenas para relacionarse con sus respectivas tablas de dimensiones. Una tabla de dimensiones se define mediante su clave primaria que sirve para definir la integridad referencial con la tabla de hechos con la que está relacionada junto con otros atributos [3]. Este tipo de estructura suele denominarse diseño en estrella.

Algunos autores como Kimball et al [3] defienden que el modelado dimensional es la única técnica viable para proporcionar datos a los usuarios finales de un almacén de datos y, además, cuenta con una serie de ventajas que no tiene el modelado E/R y su transformación a relacional.

Sin embargo, no hay pruebas sobre esta afirmación por lo que decidimos hacer una serie de experimentos con la intención de demostrar que el uso del diseño en estrella proporciona almacenes de datos más sencillos de utilizar que cuando se utiliza la forma de diseño tradicional (E/R y su transformación a relacional).

**Resumen:** en los últimos años los almacenes de datos (*Data Warehouses*) se han convertido en la tendencia más importante de la informática empresarial por lo que es fundamental que el diseño de los mismos se haga de forma que sean eficientes y sencillos de utilizar. A pesar de que está generalmente aceptado que el diseño en estrella es la mejor forma de diseñar almacenes de datos en sistemas de gestión de bases de datos relacionales, no se han realizado estudios para confirmar este supuesto. Con el objetivo de determinar si realmente el diseño en estrella consigue almacenes de datos más comprensibles, estamos llevando a cabo una serie de experimentos. En este artículo presentamos los trabajos realizados hasta este momento y que, aunque no podemos decir que hayan dado resultados definitivos, sí parecen indicar que un modelo de almacén de datos utilizando el diseño en estrella no es más difícil de entender que si se construye el almacén de datos de forma tradicional (usando el modelo E/R y luego transformándolo al modelo relacional), habiendo incluso obtenido en alguno de los experimentos que es más fácil al utilizar el diseño en estrella.

**Palabras clave:** almacenes de datos, esquemas en estrella, experimentos.

Además, al llevar a cabo este proceso experimental tuvimos en cuenta que, tal y como se indica en [4], las réplicas de los mismos son necesarias. Existen diferentes tipos de réplicas: réplicas que no varían las hipótesis y que, a su vez, pueden ser estrictas (duplican el experimento original y son necesarias para incrementar la fiabilidad en la conclusión sobre la validez del experimento) y aquellas que modifican la forma en que el experimento se realiza, réplicas que varían las hipótesis y réplicas que extienden la teoría.

El trabajo desarrollado hasta ahora se puede dividir en dos estudios, el primero de los cuales se replicó dos veces (réplica estricta). En este artículo presentamos todos los resultados que hemos obtenido de todo este trabajo experimental.

En la siguiente sección se describe en profundidad todo el proceso experimental de cada uno de los trabajos junto con sus conclusiones así como del trabajo experimental en su conjunto. La sección tercera presenta las conclusiones obtenidas y describe los trabajos futuros.

## 2. Trabajo experimental

Hasta ahora, y con el objetivo de determinar si es mejor realizar el diseño lógico de los almacenes de datos mediante la utilización o de una metodología de diseño tradicional (que empieza por el modelo E/R y sigue con la transformación al modelo relacional) o

mediante el diseño en estrella, hemos llevado a cabo cuatro experimentos clasificados en dos grupos. El primer grupo lo componen un experimento con dos réplicas (réplica estricta). El segundo grupo está compuesto por un único experimento que surgió como evolución del primero y que puede ser considerado como una réplica que modifica la forma en que el experimento se realiza.

En todos los casos las hipótesis de trabajo son las mismas, como las variables dependiente e independiente con las que se trabaja, el material experimental y la forma de ejecutar el experimento y muchas de las amenazas a la validez de los experimentos. Antes de pasar a explicar en profundidad cada uno de los dos tipos de experimentos junto con los resultados obtenidos mostraremos detalladamente los elementos comunes.

**Hipótesis.** Las hipótesis de nuestros experimentos son:

**Hipótesis Nula (H0):** no hay diferencia entre los sujetos que usan los dos tipos de esquemas (Tradicional y Estrella) con respecto al tiempo necesario para entender el esquema.

**Hipótesis Alternativa (H1):** hay diferencia entre los sujetos que usan los dos tipos de esquemas (Tradicional y Estrella) con respecto al tiempo necesario para entender el esquema.

Fuente	Suma de cuadrados	gl	Media cuadrática	F	Sig.	Potencia
TIPO ESQUEMA	10730,09	1	10730,09	0,01	0,95	0,10

Tabla 1. Resultados de ANOVA para los alumnos de doctorado de la E.S. de Informática de Ciudad Real.



El experimento se realizó en una sola sesión.



Fuente	Suma de cuadrados	gl	Media cuadrática	F	Sig.	Potencia
TIPO ESQUEMA Hipótesis	8557,55	1	8557,55	0,09	0,82	0,10

Tabla 2. Resultados de ANOVA para los alumnos de tercer curso de la E.S. de Informática de Ciudad Real.

Variables dependientes e independientes. En cuanto a las variables, en todos los casos hemos capturado la variable dependiente mediante el tiempo utilizado por cada sujeto en realizar las tareas indicadas siendo la independiente el modelo utilizado para la representación lógica del almacén de datos (tradicional o estrella).

En las siguientes dos secciones presentaremos en profundidad aquellos aspectos propios de cada uno de los dos trabajos experimentales llevados a cabo.

Material experimental y ejecución del experimento. El material experimental y la forma de llevar a cabo el experimento fueron similares en ambos trabajos. Se proporcionaron a los sujetos los esquemas (seis en el primer trabajo y doce en el segundo) en los que se incorporaba una hoja de preguntas para que indicaran como acceder y a qué tablas de cada esquema con el fin de recuperar una determinada información del almacén (visitando el mayor número de tablas posibles). También en esta hoja debían anotar el tiempo utilizado en responder a cada una de las cuestiones.

El experimento se realizó en una sola sesión y antes de realizar el experimento se dio una explicación intensiva de qué tipo de problemas debían resolver, cómo se debía contestar a las preguntas y qué material se estaba proporcionando para la realización del experimento. En ambos casos los sujetos no tenían conocimiento de los aspectos que pretendíamos estudiar ni cuales eran las hipótesis que se habían planteado.

Amenazas. Para conseguir evitar diversas amenazas a la validez del experimento, tomamos una serie de medidas:

- Los sujetos de cada experimento tenían una experiencia y unos conocimientos parecidos. Aunque trabajar con estudiantes pueda parecer poco riguroso, existen estudios que afirman que las diferencias entre alumnos y profesionales son pequeñas y los estudios con estudiantes son viables bajo ciertas condiciones [5].
- Los dominios de los diagramas eran lo suficientemente sencillos y habituales para que no existiese ningún problema a la hora de entenderlos.

- Para evitar los efectos de aprendizaje los esquemas fueron entregados a cada sujeto en un orden diferente.

- Como era la primera vez que los sujetos realizaban un experimento de este tipo, no existían efectos de persistencia

- Los sujetos estaban motivados pues los ejercicios formaban parte de los conocimientos que debían adquirir en su formación.

- No se permitió que los sujetos hablaran entre ellos durante la prueba ni que pudieran copiar los resultados unos de otros.

- Todas las dudas fueron resueltas por la persona que conducía el experimento.

### 2.1. Trabajo experimental primero

Como ya hemos mencionado este primer trabajo está compuesto por tres experimentos (un experimento original junto con dos réplicas estrictas del mismo).

Diseño experimental. Teniendo en cuenta las hipótesis realizamos un diseño experimental compuesto por tres esquemas tradicionales y tres esquemas semánticamente equivalentes diseñados mediante diagramas en estrella. Así pues, el experimento constaba de seis esquemas. Los sujetos tenían que construir unas consultas en SQL y anotar el tiempo (en segundos) que tardaban en realizarlas.

Sujetos. En el experimento original los sujetos fueron alumnos de doctorado de la Escuela Superior de Informática de Ciudad Real (UCLM). Una de las réplicas fue realizada por alumnos de tercer curso (matriculados en la asignatura de bases de datos) de esta escuela mientras que la otra se llevó a cabo con alumnos de doctorado de la Universidad de Pinar del Río (Cuba).

En todos los casos los sujetos tenían conocimientos de diseño y uso de bases de datos. Además los alumnos de doctorado de Ciudad Real y Pinar del Río tenían conocimientos de diseño y uso de los almacenes de datos pues habían recibido esta información como parte de sus estudios de doctorado.

Fuente	Suma de cuadrados	gl	Media cuadrática	F	Sig.	Potencia
TIPO ESQUEMA Hipótesis	101117,21	1	101117,21	0,02	0,90	0,10

Tabla 3. Resultados de ANOVA en alumnos de doctorado de la Univ. de Pinar del Río.

Limitaciones. Somos conscientes de algunas limitaciones asociadas a estos experimentos como el reducido número de sujetos, de objetos y la dificultad asociada a la utilización del lenguaje SQL para la especificación de los ejercicios planteados a los sujetos.

Resultados. Para nuestro experimento fijamos un valor  $\alpha = 0,1$  como forma de aumentar la potencia de las pruebas estadísticas (es decir, la probabilidad de rechazar nuestras hipótesis cuando éstas son falsas). Debido al diseño experimental y a los datos recogidos el test más adecuado es una prueba ANOVA de medidas repetidas univariante [6].

En las tablas 1, 2 y 3 se muestran los resultados obtenidos de la aplicación del estadístico a los datos recogidos del experimento. Analizando el valor de significación (Columna Sig en las tablas) con el nivel de  $\alpha = 0,1$  vemos que todos los valores son mayores que  $\alpha$  y por tanto no podemos rechazar la hipótesis de que no existe diferencia en el tiempo utilizado para responder a las cuestiones con respecto al tipo de diseño (tradicional o estrella, variable TIPO\_ESQUEMA).

Como conclusión del experimento podemos deducir que parece no haber diferencia en la comprensión de los esquemas debido al método de diseño utilizado y que por tanto daría igual diseñar un almacén de datos utilizando el modelo en estrella o de forma tradicional. Sin embargo, como estas conclusiones pueden ser debidas a que los tamaños de los esquemas utilizados en el experimento no son muy grandes o a que se ha utilizado SQL para obtener las respuestas, decidimos replicar el experimento incorporando algunos cambios que nos permitieran evitar estas limitaciones en la medida de lo posible.

### 2.2. Trabajo experimental segundo

En el segundo trabajo y en vista de los resultados obtenidos anteriormente decidimos realizar una réplica en la que, partiendo de las mismas hipótesis y trabajando con las mismas variables, incorporamos dos cambios fundamentales (ver apartado 1), en primer lugar decidimos aumentar el número de objetos con el que trabajamos y, en segundo

	Suma de cuadrados	gl	Media cuadrática	F	Sig.	Potencia
TIPO_ESQUEM/Hipótesis	154615,00E	1	154615,00E	13,904662	0,0002491	0,9600546

Tabla 4. Resultados de ANOVA para el segundo trabajo experimental.

lugar, tratamos de facilitar la labor a los sujetos a la hora de realizar los ejercicios permitiendo que utilizaran el lenguaje natural en lugar del SQL. Así pues, este nuevo experimento queda caracterizado como sigue.

**Diseño experimental.** En esta ocasión teníamos seis esquemas tradicionales y seis esquemas semánticamente equivalentes diseñados mediante diagramas en estrella, es decir, en esta ocasión cada experimento constaba de doce esquemas. Sobre cada uno de estos doce esquemas los sujetos tenían que indicar, recordamos que en lenguaje natural, los pasos necesarios para obtener una determinada información a partir del esquema del almacén de datos y anotar el tiempo (en segundos) que tardaban en realizarlas (como en el caso anterior, ya que como ya indicamos la variable dependiente no varió).

**Sujetos.** En esta ocasión contamos con dieciocho personas, alumnos de quinto curso de la Escuela Superior de Informática de Ciudad Real (UCLM) que cursaban una asignatura de Almacenamiento y Recuperación de la Información en la que todos los conceptos relacionados con los almacenes de datos habían sido vistos. Además, todos ellos habían cursado la asignatura de Bases de Datos (obligatoria de tercer curso) en la que todos los contenidos relativos al modelo relacional son tratados en profundidad.

**Limitaciones.** En este caso somos conscientes de algunas limitaciones asociadas como el reducido número de sujetos.

**Resultados.** Antes de proceder al análisis teníamos que establecer un valor para el nivel de significación. En este caso, tratando de aumentar la potencia del test, fijamos un valor  $\alpha = 0,05$ . Debido al diseño experimental y a los datos recogidos el test más

adecuado de nuevo era una prueba ANOVA de medidas repetidas univariante (SPSS, 1997).

En la tabla 4 se muestran los resultados obtenidos de la aplicación del estadístico a los datos recogidos del experimento. Analizando el valor de significación (Columna Sig en las tablas) con el nivel de  $\alpha = 0,05$  vemos que el valor obtenido del experimento es menor que  $\alpha$  y por tanto podemos rechazar la hipótesis de que no existe diferencia en el tiempo utilizado para responder a las cuestiones con respecto al tipo de diseño (tradicional o estrella, variable TIPO\_ESQUEMA).

Debido a que el valor obtenido es significativo, es necesario hacer la diferencia de medias para obtener más información. En la tabla 5 se presentan los resultados obtenidos para este test estadístico.

En función de los resultados obtenidos para la diferencia de medias, puede derivarse que cuando el almacén de datos se ha diseñado utilizando diagramas en estrella las medias son menores por lo que podría concluirse que el diseño en estrella sería más fácil de entender que el tradicional para el diseño de almacenes de datos.

**2.3. Conclusiones de todo el trabajo experimental desarrollado**

Como puede apreciarse, los resultados obtenidos no son en absoluto definitivos ya que aunque en el primer trabajo obtuvimos que ambas técnicas de modelado podrían ser apropiadas para el diseño lógico de los almacenes de datos, en el segundo trabajo el modelo en estrella parece ser más adecuado.

Esto podría hacer que nos inclináramos a pensar que, efectivamente, utilizar el diseño en estrella no sólo no es más difícil que el relacional sino que podría incluso resultar más sencillo.

Sin embargo, para poder llegar a resultados más definitivos y fiables resulta imprescindible realizar réplicas del segundo trabajo experimental con más sujetos y con diferente experiencia (por ejemplo con diseñadores de almacenes de datos). Además, sería conveniente realizar otro tipo de réplicas, por ejemplo, variando las hipótesis o extendiendo la teoría (ver sección 1).

**3. Conclusiones y trabajos futuros**

Los almacenes de datos son una de las principales tendencias empresariales en los sistemas de información puesto que ayudan en la toma de decisiones estratégicas.

Se han propuesto diversos métodos de diseño de almacenes de datos basados en los diagramas de estrella, ya que estos diagramas supuestamente aumentan la eficacia y la comprensión de los esquemas de los almacenes de datos frente a la utilización del modelado tradicional (E/R y relacional). Aunque esta afirmación es ampliamente aceptada, no se ha demostrado empíricamente que sea cierta, por lo que decidimos realizar una serie de experimentos para poder comprobarlo. Los experimentos tratan de detectar relaciones causales entre la forma de realizar el diseño lógico de un almacén de datos (tradicional o estrella) y la comprensión del mismo. Para ello hemos llevado a cabo una serie de experimentos en el que los sujetos debían realizar consultas (en SQL o en lenguaje natural) sobre un esquema lógico de almacén de datos (tradicional o en estrella). La forma de determinar la comprensión de cada uno de los esquemas es trabajando con el tiempo requerido para llevar a cabo las operaciones indicadas.

Como conclusión de nuestro estudio podemos decir que la utilización del modelo en estrella, como se preveía, no sólo no es más difícil de entender que el relacional sino que en algún caso ha resultado ser más sencillo, lo cual está de acuerdo con lo comúnmente aceptado.

Sin embargo, para poder tener resultados fiables y conclusiones definitivas es necesario que realicemos nuevos experimentos que sean o bien réplicas estrictas de los aquí presentados o bien modifiquen las hipótesis, o bien extiendan la teoría del trabajo experimental aquí presentado. Además resulta fundamental poder contar con resultados obtenidos a partir de datos reales mediante casos de estudio.

**Referencias**

[1] M. Jarke, M. Lenzerini, Y. Vassilou, P. Vassiliadis. *Fundamentals of Data Warehouses*. Ed. Springer, 2000.  
 [2] W.H. Inmon. *Building the data warehouse*. 3ª ed. Ed. Wiley, 2002.  
 [3] R. Kimball, L. Reeves, M. Ross, W. Thornthwaite. *The datawarehouse life cycle toolkit*. Ed. Wiley, 1998.  
 [4] V. Basili, F. Shull, F. Lanubile. "Building knowledge through families of experiments". *IEEE Transactions on Software Engineering*, 25 (4), 435-437, 1999.  
 [5] M. Hörst, B. Regnell, C. Wohlin. "Using students as Subjects - A Comparative Study of Students & Professionals in Lead-Time Impact Assessment". *4th Conference on Empirical Assessment & Evaluation in Software Engineering, EASE*, Keele University, UK, 2000.  
 [6] SPSS Advanced Statistics 7.5. SPSS Inc., 1997.

Número de esquema	Diseño tradicional	Diseño en estrella	Diferencia de medias
1	174,33	164,56	9,78
2	263,78	259,72	4,06
3	245,22	243,61	1,61
4	333,83	152,67	181,17
5	168,00	158,61	9,39
6	299,83	184,78	115,06
Total	247,50	193,99	53,51

Tabla 5. Resultados de la diferencia de medias para el segundo trabajo experimental.