

MÉTRICAS PARA LA CALIDAD DE BASES DE DATOS RELACIONALES

Coral Calero, Mario Piattini,
Francisco Ruiz, Macario Polo
Grupo ALARCOS
Universidad de Castilla-La Mancha



CONTENIDOS

- Introducción
- Métricas para bases de datos relacionales
- Ejemplo
- Validación empírica
- Conclusiones

INTRODUCCIÓN

- Necesidad de las métricas

“Cuando puedas medir lo que estás diciendo y expresarlo en números, sabrás algo acerca de eso; pero cuando no puedes medirlo, cuando no puedes expresarlo en números, tus conocimientos serán escasos y no satisfactorios”

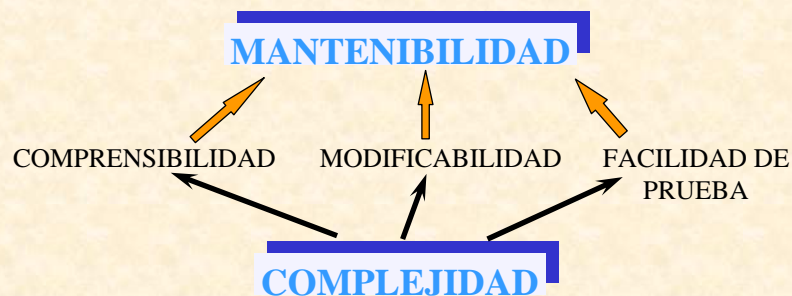
Lord Kelvin

- Métricas, también para bases de datos

3

MÉTRICAS PARA BD RELACIONALES

Li y Cheng (1987)



4

MÉTRICAS PARA BD RELACIONALES

Henderson-Sellers (1994)



5

MÉTRICAS PROPUESTAS

- Número de atributos
- Grado de referenciabilidad
- Profundidad del árbol referencial
- Ratio de normalidad
- Cohesión del esquema

$$NR = \frac{NT3NF}{NTS}$$

$$COS = \sum_{i=1}^{|US|} NTUSi^2$$

6

EJEMPLO

```

CREATE TABLE EMPLEADO
(
    NOMBREP VARCHAR(15) NOT NULL,
    INIC CHAR,
    APELLIDO VARCHAR(15) NOT NULL,
    NSS CHAR(9) NOT NULL,
    FECHAEN DATE,
    DIRECCION VARCHAR(30),
    SEXO CHAR,
    SALARIO DECIMAL(10,2),
    NSSSUPER CHAR(9),
    ND INT NOT NULL,
    CONSTRAINT CLPEMP PRIMARY KEY (NSS),
    CONSTRAINT CLESUPEREMP FOREIGN KEY (NSSSUPER) REFERENCES EMPLEADO(NSS)
        ON DELETE SET NULL ON UPDATE CASCADE,
    CONSTRAINT CLEDEPTOEMP FOREIGN KEY (ND) REFERENCES DEPARTAMENTO (NUMEROD)
        ON DELETE SET DEFAULT ON UPDATE CASCADE);

CREATE TABLE DEPARTAMENTO
(
    NOMBRED VARCHAR(15) NOT NULL,
    NUMEROD INT NOT NULL,
    NSSGTE CHAR(9) NOT NULL,
    FECHAENICGTE DATE,
    CONSTRAINT CLPDEPTO PRIMARY KEY (NUMEROD),
    CONSTRAINT CLSDEPTO UNIQUE(NOMBRED),
    CONSTRAINT CLEGTEDEPTO FOREIGN KEY (NSSGTE) REFERENCES EMPLEADO(NSS)
        ON DELETE SET DEFAULT ON UPDATE CASCADE);

CREATE TABLE LUGAR_DEPTS
(
    NUMEROD INT NOT NULL,
    LUGARD VARCHAR(15) NOT NULL,
    PRIMARY KEY (NUMEROD, LUGARD),
    FOREIGN KEY (NUMEROD) REFERENCES DEPARTAMENTO (NUMEROD)
        ON DELETE CASCADE ON UPDATE CASCADE);

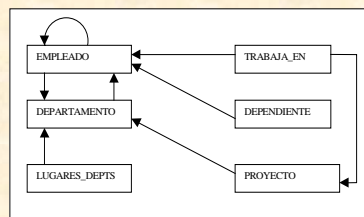
CREATE TABLE PROYECTO
(
    NOMBREP NUMEROPR VARCHAR(15) NOT NULL,
    NUMEROPR INT NOT NULL,
    LUGARPR NUMD VARCHAR(15),
    NUMD INT NOT NULL,
    PRIMARY KEY (NUMEROPR),
    UNIQUE (NOMBREP),
    FOREIGN KEY (NUMD) REFERENCES DEPARTAMENTO (NUMEROD));

CREATE TABLE TRABAJA_EN
(
    NSSE CHAR(9) NOT NULL,
    NUMP INT NOT NULL,
    HORAS DECIMAL(3,1) NOT NULL,
    PRIMARY KEY (NSSE, NUMP),
    FOREIGN KEY (NSSE) REFERENCES EMPLEADO (NSS),
    FOREIGN KEY (NUMP) REFERENCES PROYECTO (NUMEROPR));

CREATE TABLE DEPENDIENTE
(
    NSSE CHAR(9) NOT NULL,
    NOMBRE_DEPEND VARCHAR(15) NOT NULL,
    SEXO CHAR,
    FECHAAN DATE,
    RELACION VARCHAR(8),
    PRIMARY KEY (NSSE, NOMBRE_DEPEND),
    FOREIGN KEY (NSSE) REFERENCES EMPLEADO (NSS));
    
```

7

EJEMPLO



	NA	RD	DRT	NR	COS
EMPLEADO	10	2			
DEPARTAMENTO	4	1			
LUGARES_DEPTS	2	1			
PROYECTO	4	1			
TRABAJA_EN	3	2			
DEPENDIENTE	5	1			
ESQUEMA	28	8	5	1	36

8

VALIDACIÓN EMPÍRICA DE LAS MÉTRICAS

- Métricas relacionadas con la integridad referencial: RD y DRT
- Pasos de la validación: Hipótesis, Sujetos, Materiales, Diseño experimental, Resultados, Interpretación de los Resultados y Conclusiones

9

Hipótesis

- Hipótesis nula. La comprensión del esquema de la base de datos no está influida por los valores de las métricas
- Hipótesis alternativa 1. La comprensión depende del valor de DRT

10

Hipótesis

- Hipótesis alternativa 2. La comprensión depende del valor de RD
- Hipótesis alternativa 3. La comprensión depende de la combinación de los valores de DRT y de RD

11

Sujetos

- Alumnos del tercer curso de la Escuela de Informática
- Reducción de la variabilidad entre los participantes: igual curso, iguales pruebas, igual tiempo de desarrollo
- Realización previa por un grupo reducido

12

Materiales del experimento

- Se realizaron 4 tests diferentes con distintos valores para las métricas
- Material: esquema de la BD, tablas, hoja de preguntas y respuestas
- Ejercicios: inserción, eliminación y modificación de la BD

• ¿Qué tablas y cuantas columnas por tabla se verán afectadas si borramos de la tabla 5 la columna con cod1=210?

Tabla 1	Tabla 2	Tabla 3	Tabla 4	Tabla 5	Tabla 6

• ¿Qué tablas y cuantas columnas por tabla se verán afectadas si modificásemos la columna X de la fila con X=11 en la tabla 3?

Tabla 1	Tabla 2	Tabla 3	Tabla 4	Tabla 5	Tabla 6

• ¿Cuántas filas, de qué tablas y cuantas columnas por tabla habría que modificar si quisiéramos añadir una nueva fila (con todos los datos nuevos en la base de datos) a la tabla 6?

Tabla 1	Tabla 2	Tabla 3	Tabla 4	Tabla 5	Tabla 6

13

Diseño experimental

			Factor B	
			(RD)	
			BAJO	ALTO
Factor A	(TDRT)	BAJO	2,5	2,8
		ALTO	5,5	5,8

14

Resultados del experimento

Factor de variación	Q	Grados de libertad	S_i^2	F-Ratio
TDRT	18.457	1	18.5	1.67
RD	531.000	1	531	48.1
Interaction	31.339	1	31.3	2.84
Error	2560.304	1	11.0	
Total	3141.102	232		

$$F_{1,232} = 2.73$$

15

Interpretación de los resultados

☑ Hipótesis 1. Como $1.67 < 2.73$, la hipótesis es NO VÁLIDA. Por tanto DRT no afecta a los resultados

☑ Hipótesis 2. Como $48.1 > 2.73$, la hipótesis es VÁLIDA. Por tanto RD afecta a los resultados del experimento

☑ Hipótesis 3. Como $2.84 > 2.73$, la hipótesis es VÁLIDA. Por tanto la combinación de RD y DRT afecta a los resultados del experimento

16

Interpretación de los resultados

- El número de claves ajenas es un indicador sólido de la comprensibilidad de un esquema de BD
- La longitud del camino referencial por sí misma no resulta indicativa aunque sirve para modular los efectos en combinación con la otra métrica

17

CONCLUSIONES

- Métricas para bases de datos relacionales
- Verificación formal de las métricas
- Validación empírica
- Herramienta para el cálculo automático de las métricas

18