تطبيع جداول قاعدة البيانات

Normalization of Database Tables

د. عبدالناصر ضياف

سنتعلم في هذا الفصل:

- ما المقصود بشذوذ البيانات وما هي أنواعه.
- ما المقصود بـالتطبيع Normalization ومـا الـدور الـذي يلعبـه في عمليـة تصميم قاعدة البيانات
- ماهية النماذج الطبيعية: 1st NF، 2nd NF، 3rd NF، BCNF، 4th NF
- كيفية تحويل النماذج الطبيعية من مستوى منخفض إلى مستوى أعلى
- كيفيــة اسـتخدام كـل من التطـبيع ونمذجـة علاقـات الكائنـات ER تزامنيًـا للحصول على تصميم جيد لقاعدة البيانات
- كيف تتطلب بعض الحالات إلغاء التطبيع لتوليد المعلومات بكفاءة

تكرار البيانات وشذوذها
Data redundancy and Anomalies

رقم المدرب	اسم المدرب	القسم	المقرر الدراسي
123	محمد علي محمود	التحكم الآلي	مقدمة في التحكم الآلي
234	خالد عبدالكريم سالم	المواد العامة	فيزياء
345	أسامة ميلود عبدالله	الحاشب الآلي	مباديء برمجة
456	أسامة ميلود عبدالله	الحاشب الآلي	قواعد بيانات
567	علاء احميده عبدالهادي	الموجات اللاسلكية والاتصالات	اتصالات 1
678	علاء احميده عبدالهادي	الموجات اللاسلكية والاتصالات	كهرومغناطيسية

يبدو الجدول طبيعيا،،، ولكن دعونا نتفحصه،،،

تكرار البيانات وشذوذها
(متواصل)

- شذوذ التحديث Update Anomaly
  - يتمثل في حدوث اختلاف في بيانات من المفترض أن تتساوى في قيمها مما يسبب عدم تماسكها في قاعدة البيانات.
  - مثال: لتصحيح الخطأ في بيان قسم "**الحاشب الآلي**" يجب أن تُنفّذ العملية مرتين على الأقل. إذا لم يدرك من يقوم بالتحديث أن هناك تكرار لما يريد تعديله فلن تتم العملية بشكل صحيح.
- شذوذ الحذف Deletion Anomaly
  - يتمثل في الفقدان الغير مقصود لبيانات معينة نتيجة حذف بيانات أخرى
  - مثال: إذا تم حذف المقرّر الدراسي "**مقدمة في التحكم الآلي**" لعدم فاعليته في التدريس فسيتسبب ذلك في إزالة كل من المدرّب "محمد علي محمود" وقسم "التحكم الآلي" من قاعدة البيانات نهائيا
- شذوذ الإدخال Insertion Anomaly
  - يتمثل في عدم إمكانية إضافة بيانات لقاعدة البيانات لعدم توفّر بيانات أخرى
  - مثال: نفترض أن حقل المقرر الدراسي لا يقبل null ولدينا مدرب جديد تم تعيينه ولكن لم يتم تكليفه بمقررات بعد فباتالي لا يمكن إدخاله لقاعدة البيانات.

التطبيع وجداول قاعدة البيانات
Normalization and Database Tables

- حالات شذوذ *التحديث* و*الحذف* و*الإدخال* غير مرغوب فيها بتاتًا في أي قاعدة بيانات. يمكن تجنب هذه الحالات من خلال عملية التطبيع.

- التطبيع Normalization
  - عملية تقييم وتصحيح بُنى الجداول لتقليل التكرار في البيانات
    - يقلل من حالات شذوذ البيانات Data Anomalies
  - يعمل من خلال سلسلة من المراحل تسمى النماذج الطبيعية Normal Forms
    - النموذج الطبيعي الأول (1st NF) First normal form
    - النموذج الطبيعي الثاني (2nd NF) Second normal form
    - النموذج الطبيعي الثالث (3rd NF) Third normal form

التطبيع وجداول قاعدة البيانات
(متواصل)

- التطبيع Normalization (متواصل)

– 2nd NF أفضل من 1st NF، و3rd NF أفضل من 2nd NF

– معظم أغراض تصميم قواعد البيانات التجارية تحتاج النموذج الطبيعي الثالث كحد أدنى لعملية التطبيع

– أحيانا لا يكون المستوى الأعلى للتطبيع مرغوبًا

الحاجة للتطبيع
The Need for Normalization

- مثال: شركة تقوم بإدارة مشاريع بناء

– تحميل عملائها تكاليف الإدارة عن طريق فوترة الساعات المصروفة على كل عقد

– سعر ساعة العمل تعتمد على مركز الموظف الذي قام بالعمل

– يتم إعداد تقرير دوري يحتوي على معلومات كما بالجدول 5.1:

الحاجة للتطبيع
(متواصل)

TABLE 5.1 A Sample Report Layout

PROJ. NUM.	PROJECT NAME	EMPLOYEE NUMBER	EMPLOYEE NAME	JOB CLASS.	CHG/ HOUR	HOURS BILLED	TOTAL CHARGE
15	Evergreen	103	June E. Arbough	Elec. Engineer	$ 85.50	23.8	$ 2,011.10
		101	John G. News	Database Designer	$105.00	19.4	$ 2,037.00
		105	Alice K. Johnson*	Database Designer	$105.00	35.7	$ 3,748.50
		106	William Smithfield	Programmer	$ 35.75	12.6	$ 450.45
		102	David H. Senior	Systems Analyst	$ 96.75	23.8	$ 2,302.65
				Subtotal			**$10,549.70**
18	Amber Wave	114	Annelise Jones	Applications Designer	$ 48.10	25.6	$ 1,183.26
		118	James J. Frommer	General Support	$ 18.36	45.3	$ 831.71
		104	Anne K. Ramoras*	Systems Analyst	$ 96.75	32.4	$ 3,135.70
		112	Darlene M. Smithson	DSS Analyst	$ 45.95	45.0	$ 2,021.80
				Subtotal			**$ 7,172.47**
22	Rolling Tide	105	Alice K. Johnson	Database Designer	$105.00	65.7	$ 6,793.50
		104	Anne K. Ramoras	Systems Analyst	$ 96.75	48.4	$ 4,682.70
		113	Delbert K. Joenbrood*	Applications Designer	$ 48.10	23.6	$ 1,135.16
		111	Geoff B. Wabash	Clerical Support	$ 26.87	22.0	$ 591.14
		106	William Smithfield	Programmer	$ 35.75	12.8	$ 457.60
				Subtotal			**$13,660.10**
25	Starflight	107	Maria D. Alonzo	Programmer	$ 35.75	25.6	$ 879.45
		115	Travis B. Bawangi	Systems Analyst	$ 96.75	45.8	$ 4,431.15
		101	John G. News*	Database Designer	$105.00	56.3	$ 5,911.50
		114	Annelise Jones	Applications Designer	$ 48.10	33.1	$ 1,592.11
		108	Ralph B. Washington	Systems Analyst	$ 96.75	23.6	$ 2,283.30
		118	James J. Frommer	General Support	$ 18.36	30.5	$ 559.98
		112	Darlene M. Smithson	DSS Analyst	$ 45.95	41.4	$ 1,902.33
				Subtotal			**$17,559.82**
				Total			**$48,942.09**

Note: * indicates project leader.

الحاجة للتطبيع
(متواصل)

FIGURE 5.1 Tabular representation of the report format

Table name: RPT_FORMAT — Database name: Ch05_ConstructCo

PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	103	June E. Arbough	Elect. Engineer	$84.50	23.8
		101	John G. News	Database Designer	$105.00	19.4
		105	Alice K. Johnson *	Database Designer	$105.00	35.7
		106	William Smithfield	Programmer	$35.75	12.6
		102	David H. Senior	Systems Analyst	$96.75	23.8
18	Amber Wave	114	Annelise Jones	Applications Designer	$48.10	24.6
		118	James J. Frommer	General Support	$18.36	45.3
		104	Anne K. Ramoras *	Systems Analyst	$96.75	32.4
		112	Darlene M. Smithson	DSS Analyst	$45.95	44.0
22	Rolling Tide	105	Alice K. Johnson	Database Designer	$105.00	64.7
		104	Anne K. Ramoras	Systems Analyst	$96.75	48.4
		113	Delbert K. Joenbrood *	Applications Designer	$48.10	23.6
		111	Geoff B. Wabash	Clerical Support	$26.87	22.0
		106	William Smithfield	Programmer	$35.75	12.8
25	Starflight	107	Maria D. Alonzo	Programmer	$35.75	24.6
		115	Travis B. Bawangi	Systems Analyst	$96.75	45.8
		101	John G. News *	Database Designer	$105.00	56.3
		114	Annelise Jones	Applications Designer	$48.10	33.1
		108	Ralph B. Washington	Systems Analyst	$96.75	23.6
		118	James J. Frommer	General Support	$18.36	30.5
		112	Darlene M. Smithson	DSS Analyst	$45.95	41.4

الحاجة للتطبيع
(متواصل)

- بُنية مجموعة البيانات في الشكل 5.1 تعرقل معالجة البيانات بشكل جيد
- مع أن هذه البنية تبدو مناسبة للعمل واستخلاص التقارير بسهولة
- إلا أنها للأسف قد تُسفر عن تقرير بنتائج مختلفة اعتمادا على شذوذ البيانات الحاصل

أهداف عملية التطبيع ومراحلها

- أن يمثل كل جدول موضوعا واحدا فقط
- لا داعي لتخزين أي بند بيانات في أكثر من جدول واحد في غير الضرورة
- أن تعتمد كل سمات أو حقول الجدول على المفتاح الرئيسي

TABLE 5.2 Normal Forms

NORMAL FORM	CHARACTERISTIC	SECTION
First normal form (1NF)	Table format; no repeating groups and PK identified	5.3.1
Second normal form (2NF)	1NF and no partial dependencies	5.3.2
Third normal form (3NF)	2NF and no transitive dependencies	5.3.3

التحويل إلى النموذج الطبيعي الأول 1st NF
Conversion to 1st Normal Form

- يجب أن لا يحتوي الجدول العلائقي على مجموعات مكررة
- المجموعة المكررة
  - اشتقت اسمها من حقيقة أن مجموعة من المدخلات المتعددة من نفس النوع يمكن أن توجد في خلية واحدة
- تطبيع بُنية الجدول من شأنها التقليل من التكرار السلبي في البيانات
- <u>الخطوة 1:</u> التخلص من المجموعات المكررة
  - تمثيل البيانات على شكل جدول بحيث تحوي كل خلية على قيمة واحدة فقط، ولا توجد أكثر من خلية لنفس الغرض في نفس الصف
  - التخلص من القيم الخالية والتأكد من أن كل خلية تحتوي على قيمة البيانات المناسبة

التحويل إلى النموذج الطبيعي الأول 1st NF
(متواصل)

FIGURE 5.2 A table in first normal form

Table name: DATA_ORG_1NF — Database name: Ch05_ConstructCo

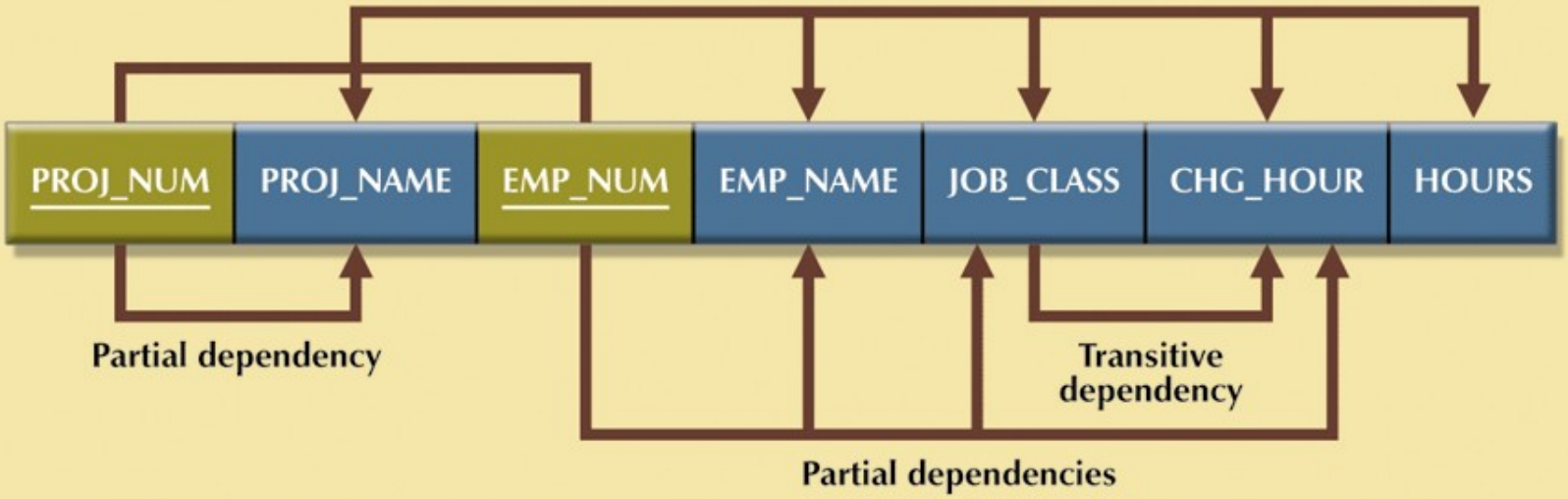
PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	103	June E. Arbough	Elect. Engineer	$84.50	23.8
15	Evergreen	101	John G. News	Database Designer	$105.00	19.4
15	Evergreen	105	Alice K. Johnson *	Database Designer	$105.00	35.7
15	Evergreen	106	William Smithfield	Programmer	$35.75	12.6
15	Evergreen	102	David H. Senior	Systems Analyst	$96.75	23.8
18	Amber Wave	114	Annelise Jones	Applications Designer	$48.10	24.6
18	Amber Wave	118	James J. Frommer	General Support	$18.36	45.3
18	Amber Wave	104	Anne K. Ramoras *	Systems Analyst	$96.75	32.4
18	Amber Wave	112	Darlene M. Smithson	DSS Analyst	$45.95	44.0
22	Rolling Tide	105	Alice K. Johnson	Database Designer	$105.00	64.7
22	Rolling Tide	104	Anne K. Ramoras	Systems Analyst	$96.75	48.4
22	Rolling Tide	113	Delbert K. Joenbrood *	Applications Designer	$48.10	23.6
22	Rolling Tide	111	Geoff B. Wabash	Clerical Support	$26.87	22.0
22	Rolling Tide	106	William Smithfield	Programmer	$35.75	12.8
25	Starflight	107	Maria D. Alonzo	Programmer	$35.75	24.6
25	Starflight	115	Travis B. Bawangi	Systems Analyst	$96.75	45.8
25	Starflight	101	John G. News *	Database Designer	$105.00	56.3
25	Starflight	114	Annelise Jones	Applications Designer	$48.10	33.1
25	Starflight	108	Ralph B. Washington	Systems Analyst	$96.75	23.6
25	Starflight	118	James J. Frommer	General Support	$18.36	30.5
25	Starflight	112	Darlene M. Smithson	DSS Analyst	$45.95	41.4

التحويل إلى النموذج الطبيعي الأول 1st NF
(متواصل)

- الخطوة 2: تحديد المفتاح الرئيسي
  - المفتاح الرئيسي يجب أن يُحدِّد قيمة الصف بشكل فردي.
  - قد نحتاج إلى تركيب المفتاح الرئيسي من أكثر من حقل
- الخطوة 3: تحديد جميع الاعتماديات
  - يمكن تصوير الاعتماديات باستخدام الرسوم التخطيطية
  - مخطط الاعتمادية (Dependency Diagram):
    - تصوّر جميع الاعتماديات الموجودة ببنية الجدول
    - تساعد في الحصول على نظرة شاملة لجميع العلاقات بين حقول الجدول
    - يجعل من غير المحتمل إغفال أي اعتمادية مهمة

التحويل إلى النموذج الطبيعي الأول 1st NF
(متواصل)

FIGURE 5.3 First normal form (1NF) dependency diagram



1NF (PROJ_NUM, EMP_NUM, PROJ_NAME, EMP_NAME, JOB_CLASS, CHG_HOURS, HOURS)

PARTIAL DEPENDENCIES:
(PROJ_NUM → PROJ_NAME)
(EMP_NUM → EMP_NAME, JOB_CLASS, CHG_HOUR)

TRANSITIVE DEPENDENCY:
(JOB CLASS → CHG_HOUR)

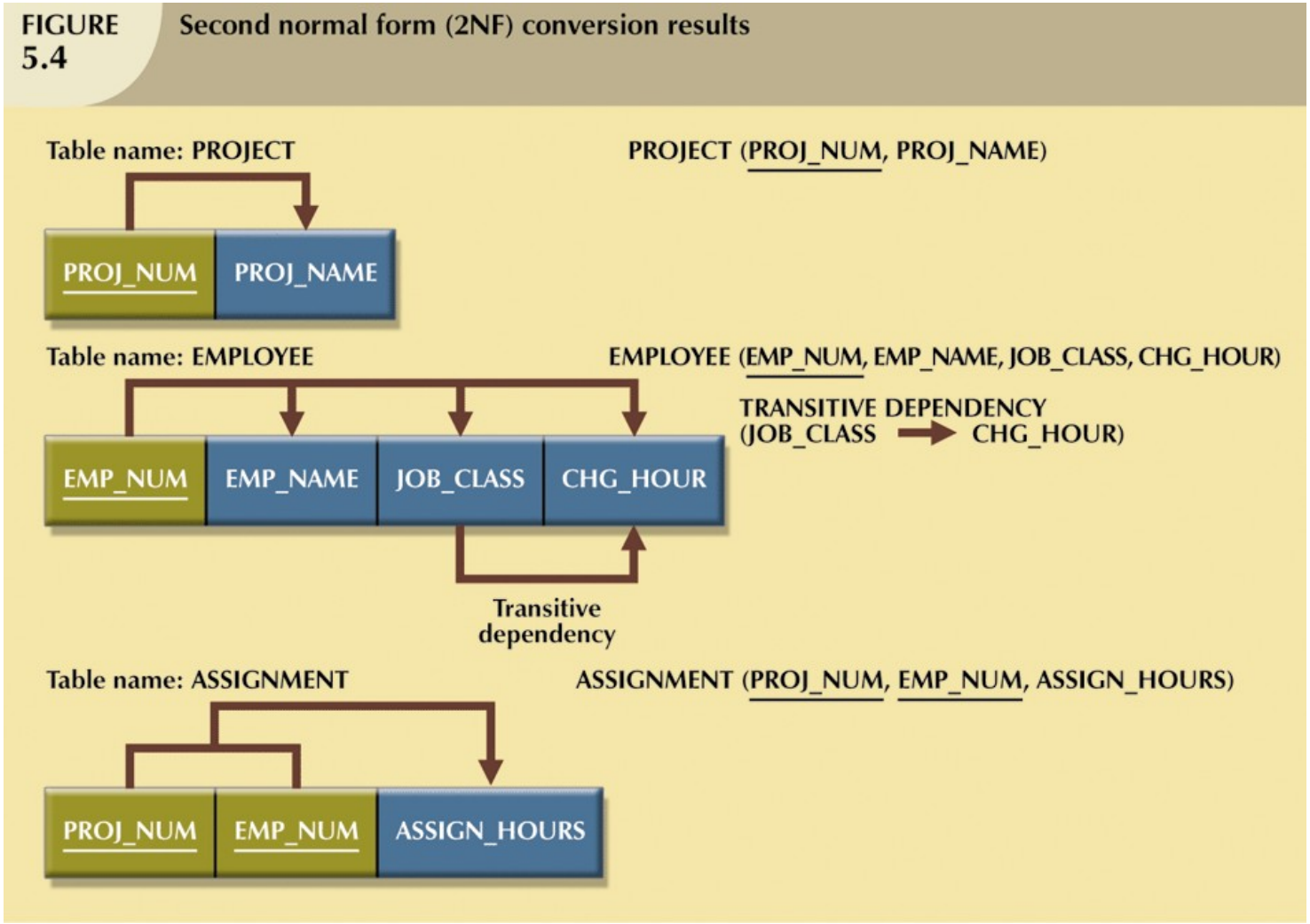
النموذج الطبيعي الأول 1st NF
(الخلاصة)

- يصف النموذج الطبيعي الأول الشكل الجدولي بحيث:
  - يتم تعريف كافة الحقول الرئيسية
  - لا توجد مجموعات مكررة في الجدول
  - تكون تبعية كل الحقول للمفتاح الرئيسي
- كل الجداول العلائقية تفي بمتطلبات الـ 1st NF
- بعض الجداول تحتوي على اعتماديات جزئية (Partial Dependencies)
  - تبعيات ترتكز على جزء فقط من المفتاح الرئيسي
  - تستخدم أحيانا لأسباب تتعلق بالأداء، ولكن ينبغي أن تستخدم بحذر
  - تبقى عرضة لتكرار البيانات

التحويل إلى النموذج الطبيعي الثاني 2nd NF
Conversion to 2nd Normal Form

- يمكن تحسين تصميم قاعدة البيانات العلائقية من خلال تحويل جداولها إلى النموذج الطبيعي الثاني بالتخلص من الاعتماديات الجزئية بها
- الخطوة 1: فصل مكونات المفتاح الرئيسي
  - كتابة كل مكون من مكونات المفتاح الرئيسي في سطر مستقل
  - ثم كتابة المفتاح المركّب الأصلي في السطر الأخير
  - سيصبح كل مكوّن مفتاحًا رئيسيا في جدول جديد
- الخطوة 2: تحديد الحقول مشتركة الاعتمادية وإلحاقها بمكوّنها الرئيسي
  - تحديد تلك الحقول التي تعتمد على حقول أخرى
  - بالوصول إلى هذه النقطة، تم التخلص من معظم حالات الشذوذ.

التحويل إلى النموذج الطبيعي الثاني 2nd NF
(متواصل)



النموذج الطبيعي الثاني 2nd NF
(الخلاصة)

- يعتبر الجدول في النموذج الطبيعي الثاني (2nd NF) عندما:
  - يحقق شروط النموذج الطبيعي الأول (1st NF)
  - لا يحتوي على أي اعتماديات جزئية
    - لا وجود لأي حقل يعتمد على جزء فقط من المفتاح الرئيسي

التحويل إلى النموذج الطبيعي الثالث 3rd NF
Conversion to 3rd Normal Form

- يمكن التخلص من حالات شذوذ البيانات بسهولة وذلك بتتبع الخطوات الثلاث التالية

- الخطوة 1: تحديد كل المحدّدات الجديدة

– كتابة محدّد كل اعتمادية متعدّية كمفتاح رئيسي لجدول جديد ملحقًا بالحقل أو الحقول المعتمدة عليه

– *المحدّد Determinant:* هو أي حقل قيمته تحدد قيم حقول أخرى في الصف.

– *الاعتمادية المتعدّية Transitive Dependency:* هي اعتمادية حقل أو أكثر على حقل آخر ليست من مكونات المفتاح الرئيسي

التحويل إلى النموذج الطبيعي الثالث 3rd NF
(متواصل)

- الخطوة 2: تحديد الحقول التابعة
  - تحديد الحقول التابعة لكل من المحدّدات المستخلصة من الخطوة 1 ثم كتابتها في جدول جديد لكل محدّد. تسمية الجداول الجديدة بمسميات تعكس محتوياتها ووظائفها
- الخطوة 3: إزالة الحقول التابعة من الاعتماديات المتعدّية
  - التخلص من جميع الحقول التابعة في العلاقات المتعدّية من كل جدول يحتوي هذا النوع من العلاقات المتعدّية
  - رسم مخطط اعتمادية جديد يبين فيه جميع الجداول التي تم استحداثها في الخطوات الثلاث 3-1
  - اختبار الجداول الجديدة والمعدّلة للتأكد من أنها لا تحتوي محدّدات أو اعتماديات غير مقبولة

التحويل إلى النموذج الطبيعي الثالث 3rd NF
(متواصل)

FIGURE 5.5 Third normal form (3NF) conversion results

Table name: PROJECT

PROJECT (PROJ_NUM, PROJ_NAME)

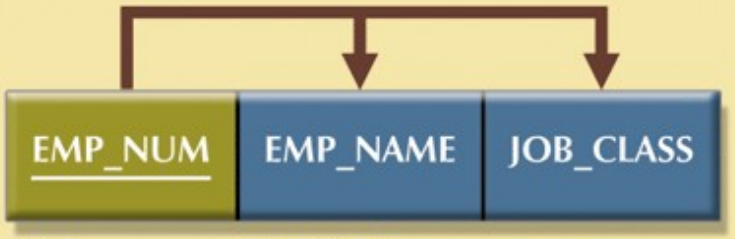


Table name: EMPLOYEE

EMPLOYEE (EMP_NUM, EMP_NAME, JOB_CLASS)

Table name: JOB

JOB (JOB_CLASS, CHG_HOUR)

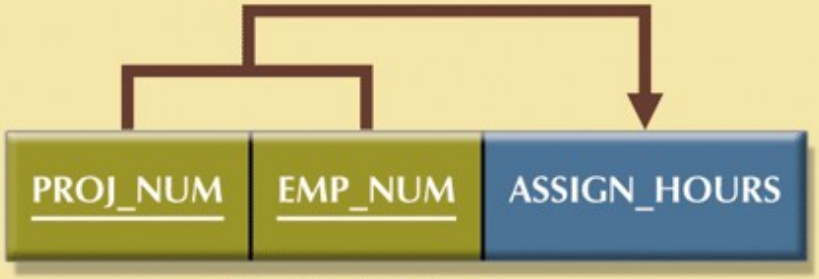


Table name: ASSIGNMENT

ASSIGNMENT (PROJ_NUM, EMP_NUM, ASSIGN_HOURS)

التحويل إلى النموذج الطبيعي الثالث 3rd NF
(الخلاصة)

- يعتبر الجدول في النموذج الطبيعي الثالث (3rd NF) عندما:
  - يحقق شروط النموذج الطبيعي الثاني (2nd NF)
  - لا يحتوي على أية اعتماديات متعدّية transitive dependency

تحسين التصميم
Improving the design

- يتم تنظيف بُنى الجداول حتى تخليصها من أي اعتماديات جزئية ومتعدّية.
- لا يمكن الاتكال فقط على عملية التطبيع للحصول على تصميم جيد.
- مع التأكيد على فاعليتها في المساعدة على التخلص من شذوذ البيانات وتجنب تكرارها.