

مبادئ قواعد المعطيات

المحاضرة الحادية عشر

د. سمر الحلبي

الصيغ المعيارية (التسوية)

Normalization

مقدمة

إن عملية وضع تصميم قاعدة البيانات في الصيغة المعيارية يشكل لبنة أساسية في عملية التصميم الجيد لقاعدة البيانات. وتم هذه العملية على عدة مراحل يتم خلالها تخلص قاعدة البيانات من التكرار غير المسوغ للبيانات بالاعتماد على قوانين الاستنتاج والاعتمادية الوظيفية. وسنقوم في هذا الدرس بالتعرف على الشروط والقوانين الالزمة للوصول بقاعدة البيانات إلى المستوى المعياري الثالث (Third Normal Form 3NF).

مشاكل تكرار البيانات : (Data Anomalies)

Employee department						
Empno	Ename	Job	Salary	Deptno	Dname	Loc
101	Sami	Clerk	3000	10	accounting	Homs
205	Khald	Manager	2500	10	accounting	Homs
203	Ali	Sales man	1200	20	Sales	Hama
502	Saed	Sales man	2100	20	Sales	Hama
606	salem	clerk	1000	30	operation	damasus

نلاحظ في الجدول السابق أن معلومات الموظف والقسم الذي يعمل فيه موجودة في جدول واحد. ونتيجة ذلك تكرار بعض البيانات مثل اسم وموقع القسم في كل سجل وهذا يؤدي إلى عدة مشاكل :

١ - **مشكلة الإضافة** : أي إننا لا نستطيع أن نضيف قسماً جديداً إلا إذا كان القسم يحتوي على، موظف ، لأن المفتاح الرئيس للجدول هو رقم الموظف.

٢ - **مشكلة التعديل** : نلاحظ تكرار اسم وموقع القسم فإذا قمنا بتعديل موقع (LOC) القسم رقم ٢٠ من Hama إلى Homs فلابد من إجراء عملية التعديل لجميع الموظفين في هذا القسم وإلا ستؤدي هذه العملية إلى عدم توافقية البيانات أي نفس رقم القسم ولتكن أكثر من موقع . وكذلك إذا تمت عملية التغيير عند الموظف رقم ٢٠٢ عن طريق الخطأ . وبالتالي لو قمنا بعملية استرجاع لجميع الموظفين الذين يعملون في Hama فإن الموظف رقم ٢٠٢ لن يظهر بين الموظفين .

٣ - **مشكلة الحذف** : نلاحظ أن القسم رقم ٢٠ يحتوي على موظف واحد فقط ، فلو قمنا بحذف الموظف رقم ٦٠٦ فإن معلومات القسم رقم ٢٠ سوف تختفي من الجدول .

الاعتمادية الوظيفية (Functional Dependency FD) :

وهي اعتماد قيمة إحدى صفات الكيان على قيمة صفة (صفات) أخرى ويرمز لها بالرمز (\rightarrow)

$$A \rightarrow B$$

مثال

يعني أن B تعتمد اعتماداً وظيفياً على A وهنا نستطيع أن نقول أن قيمة A تحدد قيمة B . ومن خلال تحديد الاعتمادية نستطيع أن نحدد المكان الذي يجب أن توضع فيه الصفة وهذا وبالتالي يؤدي إلى وضع البيانات في المكان الصحيح ونخلص من عملية تكرار البيانات وما يتترتب على التكرار من مشاكل (Anomalies) .

مثال: لـ كل موظف اسم واحد فتحل ولـ كل موظف قسم واحد ي عمل فيه إذا :

$$FD1 : Empno \rightarrow Ename$$

$$FD2 : Empno \rightarrow Deptno$$

ويمكن أن نعيد كتابة هذه الاعتمادية على الشكل التالي

$$FD1 : Empno \rightarrow Ename,Deptno$$

FD : Functional Dependency

قواعد الاستنتاج

وهي عبارة عن مجموعة من القواعد تستخدم في عملية تحديد الاعتمادية الوظيفية (Functional Dependency FD) وتتلخص هذه القواعد بستة قواعد على النحو التالي :

١ - الانعكاسية **Reflexive** : إذا كانت Y جزء من X (Y محتواه في X)

فإن X تحدد Y

$$1- X \sqsubseteq Y : X \rightarrow Y$$

٢ - قاعدة الزيادة أو الإضافة **Augmentation** : إذا كان X تحدد Y فإن إضافة Z إلى X

تعني أنه بالإمكان إضافة Z إلى Y

$$2- \{X \rightarrow Y\} \models XZ \rightarrow YZ$$

٣ - قاعدة التعدي **Transitive** : تعني أنه إذا كانت X تحدد Y وكانت Y تحدد Z

فإن X تحدد Z .

$$3- \{X \rightarrow Y, Y \rightarrow Z\} \models X \rightarrow Z$$

٤ - قاعدة الاتحاد **Union** : تعني أنه إذا كانت X تحدد Y و X تحدد Z فإننا نستطيع أن

نقول أن X تحدد YZ .

قواعد الاستنتاج:

$$4- \{X \rightarrow Y, X \rightarrow Z\} \models X \rightarrow YZ$$

٥ - قاعدة التقسيم Decomposition وهي عكس قاعدة الاتحداد

$$5- \{X \rightarrow YZ\} \models X \rightarrow Y, X \rightarrow Z$$

٦ - قاعدة التعدي الزائف Pseudotransitive تشبه قاعدة التعدي

$$6- \{X \rightarrow Y, WY \rightarrow Z\} \models WX \rightarrow Y$$

\models تعني أنه إذا تحقق الطرف الأيسر فإننا نستطيع استنتاج الطرف الأيمن .

تعريف الصيغة المعيارية الأولى (First Normal Form 1NF):

نستطيع أن نقول أن الجدول في الصيغة المعيارية الأولى إذا كانت جميع أعمدة الجدول تحتوي على بيانات بسيطة أو مفردة (غير مركبة) أي إن كل عمود يحتوي على قيمة واحدة فقط.

مثال ١ يمثل الجدول التالي معلومات موظف Employee:

No	name			adresse		
	fname	mid	lname	city	street	House no
100	ali	Salem	musa	homs	hamidia	210
120	Saed	eisa	Ali	homs	nezha	202

تعريف الصيغة المعيارية الأولى (First Normal Form 1NF) :

نلاحظ في الجدول أن الاسم يتكون من ثلاثة أجزاء وكذلك العنوان وبالتالي لا نستطيع أن نخزن قيمة واحدة في عمود الاسم وكذلك بالنسبة للعنوان وهذا يخالف شروط قاعدة البيانات بأن العمود يجب أن يحتوي على قيمة واحدة فقط. وهذا يعني أن الجدول السابق لا ينطبق عليه شرط الصيغة المعيارية الأولى، **1NF**. ولوضع الجدول في الصيغة المعيارية الأولى يجب تقسيم الأعمدة المركبة إلى أعمدة بسيطة.

No	fname	mid	lname	city	street	House no
100	ali	Salem	musa	homs	hamidia	210
120	Saed	eisa	Ali	homs	nezha	202

لقد قمنا بتقسيم الأعمدة المركبة إلى أعمدة بسيطة وبالتالي نستطيع أن نقول أن الجدول الآن في الصيغة المعيارية الأولى . **1NF**.

مثال ٢ : يمثل الجدول التالي سجل ساعات العمل HOURS لموظفي في عدد من المشاريع
والقسم الذي يشرف على تنفيذ المشروع PROJECTS

NO	Name	Project_Code	Hours	Deptno	Dname
210	Ali	P1,p2,p3	12,20,40	10,20,20	Research, Operation, Operation
201	Salem	P1,p3	30,15	10,20	Research Operation
305	Ali	P2,p3	40,20	20,20	Operation, Operation

كما هو مبين في الجدول السابق فإن هناك عدداً من الأعمدة تحتوي على أكثر من قيمة مثل رمز المشروع Project_Code وكذلك عدد ساعات العمل Hours والأقسام Deptno التي تشرف على المشاريع . وهذا يعني أن الجدول ليس في الصيغة المعيارية الأولى ، ولتحويله يجب أن نقوم بتقسيم الجدول على النحو التالي للتخلص من هذه المشكلة.

NO	Name	Project_Code	Hours	Deptno	Dname
210	Ali	P1	12	10	Research
210	Ali	P2	20	20	Operation
210	Ali	P3	40	20	Operation
201	Salem	P1	30	10	Research
201	Salem	P3	15	20	Operation
305	Ali	P2	40	20	Operation
305	ali	P3	20	20	Operation

ولكن تبرز هنا لدينا مشكلة جديدة وهي إيجاد مفتاح رئيسي للجدول إذ أصبح رقم الموظف لا يصلح لأن يكون مفتاحا رئيسا للجدول (Primary Key) لأن من شروط المفتاح الرئيسي أن لا يتكرر في أكثر من صف لنقوم الآن باستخدام الاعتمادية الوظيفية لمحاولة إيجاد المفتاح الرئيس للجدول

FD 1 : No → Name

حيث إن لكل موظف اسم واحد .

FD 2 : Project_Code → Deptno

حيث إن لكل مشروع قسم واحد يشرف عليه .

FD 3 : Deptno → Dname

حيث إن لكل قسم اسم واحد .

أما بالنسبة لبقية العناصر فمثلاً اسم الموظف لا يحدد شيئاً لأنه يوجد هناك أكثر من موظف اسمه Ali فالاسم لا يحدد الرقم وكذلك فإن علي يعمل في أكثر من مشروع .

وكذلك رمز المشروع لا يحدد عدد الساعات ولا الموظفين الذين يعملون فيه فالمشروع P1 يعمل فيه أكثر من موظف وبساعات مختلفة .

أما بالنسبة للقسم فلا يحدد الموظفين ولا المشاريع فمثلاً القسم ٢٠ يشرف على أكثر من مشروع هذه المشاريع يعمل فيها أكثر من موظف .

ففي هذه الحالة يجب علينا القيام بمحاولة جديدة لإيجاد المفتاح الرئيس وتتلخص هذه العملية بمحاولة إيجاد مفتاح مركب (تركيب أكثر من صفة لتشكل المفتاح الرئيس) يقوم بتحديد جميع الصفات في الجدول :

سنقوم بأخذ رقم الموظف مع رقم المشروع

FD 4 :No, Project_Code → name

FD 5 :No, Project_Code → Deptno

FD 6 :No, Project_Code → Hours

FD 7 : Deptno → Dname

FD 8 :No, Project_Code → Name ,Hours, Deptno, Dname

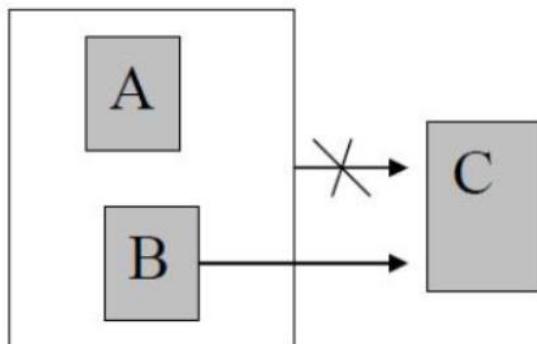
FD4,FD5 تطبق من FD1,FD2 حيث إن رقم الموظف وحدة يحدد الاسم وكذلك رمز المشروع يحدد القسم ، أما بالنسبة ل FD5 فإنها تطبق لأن رقم الموظف ورمز المشروع يحدان عمل الموظف في ذلك المشروع ، وبالتالي نكون قد حصلنا على مفتاح رئيس لهذا الجدول وكذلك قمنا بوضعه في الصيغة المعيارية الأولى (1NF).

تعريف الصيغة المعيارية الثانية (Second Normal Form 2NF)

نستطيع أن نقول أن الجدول في الصيغة المعيارية الثانية:

- ١ - إذا كان الجدول في الصيغة المعيارية الأولى.
- ٢ - إذا لم يحتوي الجدول على اعتمادية جزئية.

الاعتمادية الجزئية: هي أن تعتمد بعض الأعمدة (الصفات) اعتماداً وظيفياً على جزء من المفتاح الرئيس



نلاحظ أن A, B تحدد C أي إن C تعتمد اعتماداً وظيفياً على A, B وكذلك أن B تحدد C أي إن C تعتمد اعتماداً وظيفياً على B .

مثال

هذا الجدول يحتوى على اعتمادية جزئية .

NO	Project_Code	Hours
210	P1	12
210	P2	20
210	P3	40
201	P1	30
201	P3	15
305	P2	40
305	P3	20

مثال

والآن هل الجدول السابق في الصيغة المعيارية الثانية ؟

وللإجابة على ذلك نجيب على السؤالين التاليين :

١ - هل الجدول في الصيغة المعيارية الأولى ؟

نعم، لأنه لا توجد هناك قيم متكررة ، كل عمود يحتوي على قيمة واحدة فقط .

٢ - هل توجد هناك اعتمادية جزئية ؟

ولمعرفة ذلك يجب أن نحدد الاعتمادية الوظيفية

FD 1 : No → Name

FD 2 : Project_Code → Deptno, Dname

FD 3 : No, Project_Code → name , deptno, hours

المفتاح الرئيس هو Project_Code ، ولكن No يحدد Name إذا هناك اعتمادية جزئية وكذلك Deptno يحدد Project_Code و هذه اعتمادية جزئية أخرى .

مثال

وللتخلص من هذه المشكلة يجب أن تقوم بتقسيم الجدول إلى جداول بحيث يضم كل منها الجزء من المفتاح والأعمدة التي تعتمد عليه ونبقي فتح المفتاح المركب مع الأعمدة التي تعتمد عليه:

- ١ - تقوم بنقل اسم ورقم الموظف إلى جدول جديد ونبقي نسخة من رقم الموظف في الجدول الأصلي لأنه جزء من المفتاح الرئيس).
- ٢ - تقوم بنقل رمز المشروع ورقم القسم إلى جدول جديد ونبقي نسخة رمز المشروع في الجدول الأصلي (لأنه جزء من المفتاح الرئيس).
- ٣ - نبقي بقية الأعمدة كما هي (عدد الساعات).
- ٤ - وبالتالي تصبح الجداول على النحو التالي بعد عملية التقسيم :

مثال

NO	Name
210	Ali
210	Ali
210	Ali
201	Salem
201	Salem
305	Ali
305	ali

Project_Code	Deptno	Dname
P1	10	Research
P2	20	Operation
P3	20	Operation

مثال

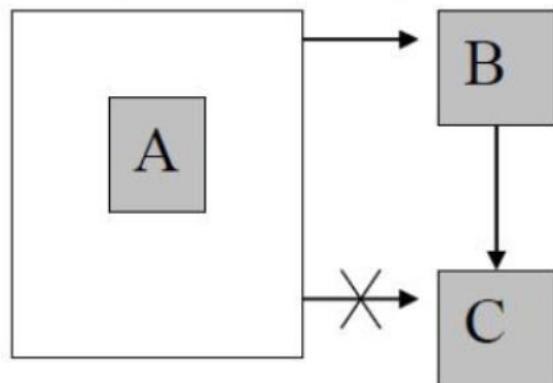
NO	Project_Code	Hours
210	P1	12
210	P2	20
210	P3	40
201	P1	30
201	P3	15
305	P2	40
305	P3	20

تعريف الصيغة المعيارية الثالثة (Third Normal Form 3NF) :

نستطيع أن نقول أن الجدول في الصيغة المعيارية الثالثة:

- ١ - إذا كان الجدول في الصيغة المعيارية الثانية.
- ٢ - إذا لم يحتوي الجدول على اعتمادية متعددة.

العتمادية المتعددة: هي أن تعتمد بعض الأعمدة (الصفات) اعتماداً وظيفياً على صفة غير المفتاح الرئيس.



نلاحظ أن A تحدد B,C أي إن C تعتمد اعتماداً وظيفياً على A وكذلك أن B تحدد C أي إن C تعتمد اعتماداً وظيفياً على B. وفي حالة نستطيع أن نقول أن هذا الجدول يحتوي على اعتمادية متعددة.

مثال

والآن هل الجداول السابقة في الصيغة المعيارية الثالثة ؟

وللإجابة على ذلك نجيب على السؤالين التاليين :

١ - هل الجداول في الصيغة المعيارية الثانية ؟

نلاحظ أن جميع الجداول في الصيغة المعيارية الثانية حيث لا يوجد فيها اعتمادية جزئية .

٢ - هل توجد هناك اعتمادية متعددة ؟

ولمعرفة ذلك يجب أن نحدد الاعتمادية الوظيفية لـ كل جدول

أ - الجدول الأول

FD 1 : No → Name

لا توجد اعتمادية متعددة .

مثال

ب - الجدول الثاني

FD 1 :No, Project_Code → hours
لا توجد اعتمادية متعددة.

ج - الجدول الثالث

FD 1 : Project_Code → Deptno,Dname
FD 2 : Deptno → Dname

المفتاح الرئيس هو Project_Code يحدد Deptno و Dname في نفس الوقت فإن Deptno يحدد Dname أي إن هناك اعتمادية متعددة . وللتخلص من هذه المشكلة يجب أن تقوم ب三分ي الجدول إلى جداول بحيث يضم كل منها الأعمدة التي تعتمد على بعض ونبقى المفتاح مع الأعمدة التي تعتمد عليه وحدة فتحل مع إبقاء المحدد الجديد (Deptno).

١ - نقوم بنقل رقم و اسم القسم إلى جدول جديد ونبقى نسخة من رقم القسم في الجدول الأصلي.

مثال

وبالتالي تصبح الجداول على النحو التالي بعد عملية التقسيم :

NO	Project_Code	Hours
210	P1	12
210	P2	20
210	P3	40
201	P1	30
201	P3	15
305	P2	40
305	P3	20

NO	Name
210	Ali
210	Ali
210	Ali
201	Salem
201	Salem
305	Ali
305	ali

مثال

Project_Code	Deptno
P1	10
P2	20
P3	20

Deptno	Dname
10	Research
20	Operation

الآن نستطيع أن نقول أن هذه الجداول هي في الصيغة المعيارية الثالثة **3NF** وتعتبر هذه الصيغة مقبولة
معظم مصممي قواعد البيانات .

تمرين 1

شركة تتلقى طلبات العملاء بالبريد وتقوم ب تخزين البيانات التالية لكل طلب: رقم الطلب، تاريخ الطلب، رقم العميل، اسم العميل، عنوان العميل، هاتف العميل، رقم كل صنف بالطلب، وصف الصنف، سعر الصنف ،الكمية المطلوبة من الصنف .

المطلوب : حول علاقة الطلبات التالية إلى الشكل النظامي الثالث 3NF.

ORDER(Order Number, Date, Cust_Number, Name, Address, Phone,
Item Number, Description, Price, Qty)

الطلبات (رقم الطلب، تاريخ الطلب، رقم العميل، اسم العميل، عنوان العميل، هاتف العميل، رقم الصنف، وصف الصنف، سعر الصنف، كمية الصنف)

ORDER(Order Number, Date, Cust_Number, Name, Address, Phone,
Item Number, Description, Price, Qty)

1NF

نجعل تقاطع كل سطر مع عمود يملك قيمة واحدة

2NF

-1-

LINEITEM (Order Number, Item Number, Description, Price, Qty)

ORDER (Order Number, Date, Cust_Number, Name, Address, Phone)

-2-

LINEITEM (Order Number, Item Number, Qty)

ITEM (Item Number, Description, Price)

ORDER (Order Number, Date, Cust_Number, Name, Address, Phone)

3NF

LINEITEM (Order Number, Item Number, Qty)

ITEM (Item Number, Description, Price)

ORDER (Order Number, Date, Cust_Number)

CUSTOMER (Cust Number, Name, Address, Phone)

تمرين 2

لدى مريض سلسلة من المواجهات مع استشاريين(consultants) في العديد من المستشفيات. يشرف على علاج المريض ممارس عام (General Practitioner GP). في يوم ما يحضر الاستشاري إلى مشفى واحد ويكون في هذا اليوم لدى المريض موعد واحد (Appointment) مع استشاري معين. أسماء الاستشاريين وأسماء المستشفيات فريدة ولا تكرر بينما أسماء المرضى وأسماء الممارس العام من الممكن تكرارها. كل استشاري له رقم هاتف وحيد.

المطلوب : حول علاقة المرضى التالية إلى الشكل النظامي الثالث 3NF.

PATIENT(PID, Name, GP_Number, GPName, GPAddress, AppointmentDate,
AppointmentTime, ConsultantName, ConsultantPhone, Hospital,
HospitalAddress)

مرضى (رقم المريض،اسم المريض،رقم الممارس،اسم الممارس،عنوان الممارس، تاريخ الكشف، وقت الكشف، اسم الاستشاري، رقم الهاتف، اسم المشفى، عنوان المشفى)

PATIENT(PID, Name, GP_Number, GPName, GPAddress, AppointmentDate,
AppointmentTime, ConsultantName, ConsultantPhone, Hospital,
HospitalAddress)

1NF

تمرين 2

نجعل تقاطع كل سطر مع عمود يملك قيمة واحدة

2NF

APPOINT (PID, AppointmentDate, AppointmentTime, ConsultantName,
ConsultantPhone, Hospital, HospitalAddress)

PATIENT (PID, Name, GP_Number, GPName, GPAddress)

3NF

APPOINT (PID, AppointmentDate, AppointmentTime, ConsultantName, Hospital)

CONS (ConsultantName, ConsultantPhone)

HOSP (Hospital, HospitalAddress)

PATIENT (PID, Name, GP_Number)

GP (GP_Number, GPName, GPAddress)