

# Бази от данни

Релационен модел на данни. Нормализация.

# Дизайн на БД. Аномалии

- Когато създаваме релационен модел на БД, може да се получат следните аномалии:
  - Излишества когато се повтаря информацията без да е необходимо
  - Аномалии при обновяване когато направим промени в един кортеж, а забравим да направим същите промени в свързаните с него кортежи
  - Аномалии при изтриване когато изтрием даден ред и това доведе до загуба на информация
  - Например ако имаме актьор, който участва във филм и ние искаме да изтрием актьора, а като ефект да получим и изтрита информация за филма, в който е участвал актьора.

StarMovies(title, year, length, filtype, studioName, starName)

# Декомпозиране на релации

- Начин да избегнем излишества е чрез декомпозиране на релации.
- Декомпозирането на една релация R означава разделяне на атрибутите на релацията R в две нови релации
- ▶ Например: Нека ни е дадена релацията R(AI,A2, ..,An), можем да декомпозираме релацията R до две релации S(BI, B2,..., Bm) и T(CI, C2, .., Ck), такива че:
  - ► {AI,A2, ...,An}={BI, B2, .., Bm} U {CI, C2, .., Ck}
  - Кортежите в релацията S са проекция по атрибутите В I, В2,
    ..., Вт на релацията R, а тези на релацията Т по атрибутите С I, С2, ..., Ск на релацията R

# Нормална форма на Бойс-Код

- Целта на декомпозицията е да замени релацията с други, които да не позволяват аномалии. Има просто правило, което гарантира, че ако една релация е подчинена на него то в съответната релация не може да възникнат аномалии.
- ▶ Казваме, че една релация се намира в Нормална форма на Бойс-Код (НФБК), тогава и само тогава когато ако А1,А2, ..An->В е нетривиална Ф3, която е в сила за R, то {A1,A2, ..An} е суперключ за R
- Или казано по друг начин лявата страна на всяка нетривиална
  Ф3, която е в сила за R трябва да съдържа ключ на релацията R
- Ако се случи, че има Ф3, които са в сила за релацията но не отговарят на посоченото по-горе правило (НФБК), тогава трябва да декомпозираме за да приведем релацията до НФБК

# Нормална форма на Бойс-Код

- Пример: Релацията
  StarMovies(title, year, length, filtype, studioName, starName)
- Не е в НФБК
- За да проверим това, нека първо да определим кой е ключът на релацията
  - Ключът e {title, year, starname}
  - Тогава всяко множество от атрибути, което съдържа тези три атрибута е супер-ключ
  - За релацията е в сила Ф3:
    - title, year > length, filmtype, studioName
    - ▶ Също така знаем че title, year не определя функционално starname
    - Ако релацията беше в НФБК, то в лявата част на ФЗ щеше да бъде сперключ
    - В случая това не е така

# Нормална форма на Бойс-Код

- Пример2: Релацията
  Movies1(title, year, length, filtype, studioName)
- Тя е в НФБК
- За нея е в сила ФЗ

title, year - > length, filmtype, studioName

 Лявата част съдържа ключа, следователно релацията е в НФБК

#### Твърдение

- Всяка релация, която има само два атрибута се намира в НФБК
- Доказателство:
- Нека R(A,B)
  - Ако нямаме нетривиални Ф3, т.е. всички Ф3 които се в сила за релацията са тривиални, тогава релацията се намира в НФБК, тъй като само не тривиалните Ф3 могат да нарушат правилото на НФБК. Единственият ключ за релацията ще бъде {A,B}
  - В случай че А->В, но не е в сила че В->А, тогава А е единственият ключ за релацията и всички нетривиални ФЗ, които са в сила за релацията го съдържат в лявата си част, т.е. релацията е в НФБК
  - Аналогично ако B->A, но не е в сила че A->B
  - В случай че A->В и В->А, тогава и атрибута А и атрибута В могат да бъдат ключ за релацията. Всички Ф3 съдържат ключ в лявата си част на правилото, следователно релацията е в НФБК

### Алгоритъм за свеждане към НФБК

- Първо се търсят нетривиални Ф3: AI,A2, ..,An->BI, B2,..,Вm такива че да нарушават НФБК, т.е. лявата част да не съдържа ключа
- По правилото за комбиниране добавяме в дясната част всички атрибути, който са функционално определени от AI,A2, ..,An
- Правим декомпозиция на релацията на две релации.
  - Първата релация съдържа атрибутите от лявата част на функционалната зависимост и тези атрибути, които не са функционално определени от тях
  - Втората релация съдържа атрибутите от лявата част на ФЗ и атрибутите от дясната част

#### Алгоритъм за свеждане към НФБК - Пример

StarMovies(title, year, length, filtype, studioName, starName)

Ф3, която нарушава правилото е

title, year - > length, filmtype, studioName

- По горе-споменатият алгоритъм разделяме релацията на две релации
  - Movies1(title, year, length, filtype, studioName)
  - StarMovies(title, year, starname)

#### Нормализация

- **Нормализация** наричаме процеса на декомпозиране на релация на няколко релации с цел отстраняване на аномалии и повторения.
- **Ф3:** Нека R(A, B, C) е релация. Функционална зависимост (Ф3) за релацията R, наричаме твърдение от вида A, B -> C

### Нормални форми

- ▶ Първа нормална форма (1НФ) изисква всеки компонент в кортежите на релацията да има атомарна стройност.
- ▶ **Втора нормална форма (2НФ)** изисква релацията да е в 1НФ и всеки атрибут да е функционално зависим от атрибутите, съставляващи първичния ключ, но не и от негово подмножество.
- ▶ Трета нормална форма (3НФ) изисква релацията да е в 2НФ и ако АІ, А2, ..Ап->В е нетривиална ФЗ която е в сила за R, то или {АІ,А2, ..Ап} да е супер-ключ за R или В да е част от ключ.
- ▶ Нормална форма на Бойс-Код (НФБК) изисква релацията да е в ЗНФ и само ако AI,A2, ..An->В е нетривиална ФЗ, която е в сила за R, то {AI, A2, ..An} да е супер-ключ за R.

- ▶ Да се направи дизайн на БД, съдържаща информация за поръчки. Известно ни е, че в БД ще се пази информация за дата на поръчката, номер на продукт – уникален за всеки продукт, име на продукт, цена на продукт, ДДС за съответния продукт, номер на клиент – уникален за всеки клиент, име на клиент, количество, което клиентът е поръчал от продукта в съответната поръчка, сума на поръчката без ДДС, сума на поръчката с ДДС.
- В сила са следните предположения:
  - Стойността на ДДС за продукта може да варира за различните продукти.
  - Поръчките на клиентите, направени в един и същ ден (на една и съща дата) се обединяват, т.е. има само една поръчка за клиент за дадена дата.
  - сума на поръчката с ДДС = сума на поръчката без ДДС + ДДС
- 1. Направете модел на БД
- 2. Определете ключовете за релациите (релацията).

#### Решение

**Подход 2:** Поставяме всички атрибути (характеристики) на данните в една табица. Намираме всички ФЗ за тази релация. Нормализираме.

1. Поставяме всичко в една релация

Orders(pld, pName, pPrice, Amount, cld, cName, oDate, VAT, nTotal, gTotal)

- 2. Намираме всички Ф3 за релацията Order (от дадената ни информация)
  - ▶ pId -> pName, pPrice, VAT (Ф3-1)
  - $\triangleright$  cId  $\rightarrow$  cName ( $\Phi$ 3-2)
  - $\triangleright$  pId, cId, oDate -> Amount ( $\Phi$ 3-3)
  - pPrice, Amount -> nTotal
  - ▶ nTotal, VAT-> gTotal
- 3. Намираме ключа за Order: {pld, cld, oDate}

### Решение – Подход 2 (Нормализация)

#### 4. Декомпозираме до НФБК:

- Намираме Ф3, която нарушава НФБК (Ф3-1, Ф3-2)
- Декомпозираме до две релации Orders I и Orders2 (по алгоритъма за привеждане до НФБК)
- 3. За Orders2 повтаряме стъпка 1 и 2, докато стигнем до нормализиран модел.

#### Решение:

- Orders(pld, pName, pPrice, Amount, cld, cName, oDate, VAT, nTotal, gTotal) и
  Ф3-1, следва:
  - Orders I (<u>pld</u>, pName, pPrice, VAT) = Products(<u>pld</u>, pName, pPrice, VAT)
  - Orders2(<u>cID</u>, cName, <u>pID</u>, <u>oDate</u>, Amount, nTotal, gTotal)

# Решение – Подход 2 (Нормализация)

- Products(pld, pName, pPrice, VAT)
- Orders2(clD, cName, plD, oDate, Amount, nTotal, gTotal) и Ф3-2 =>
  - Orders3(<u>cID</u>, cName) = Customers(<u>cID</u>, cName)
  - Orders4(pID, cID, oDate, Amount, nTotal, gTotal)

#### • Получаваме:

- ▶ Products(<u>pld</u>, pName, pPrice, VAT) НФБК
- ► Customers(<u>cID</u>, cName) НФБК
- Orders(<u>pID</u>, <u>cID</u>, <u>oDate</u>, Amount, nTotal, gTotal) НФБК



Книга: Database systems. A practical approach to design, implementation and management, Автори: Thomas Connolly, Carolyn Begg

# Ненормализиран модел

#### ClientRental

clientNo	cName	propertyNo	pAddress	rentStart	rentFinish	rent	ownerNo	oName
CR76	John Kay	PG4	6 Lawrence St, Glasgow	1-Jul-12	31-Aug-13	350	CO40	Tina Murphy
		PG16	5 Novar Dr, Glasgow	1-Sep-13	1-Sep-14	50	CO93	Tony Shaw
CR56	Aline Stewart	PG4	6 Lawrence St, Glasgow	1-Sep-11	10-June-12	350	CO40	Tina Murphy
		PG36	2 Manor Rd, Glasgow	10-Oct-12	1-Dec-13	375	CO93	Tony Shaw
		PG16	5 Novar Dr, Glasgow	1-Nov-14	10-Aug-15	450	CO93	Tony Shaw

Figure 14.10 ClientRental unnormalized table.

#### 1НФ

**Първа нормална форма (1НФ)** – изисква всеки компонент в кортежите да е атомарна стройност.

#### ClientRental

clientNo	propertyNo	cName	pAddress	rentStart	rentFinish	rent	ownerNo	oName
CR76	PG4	John Kay	6 Lawrence St, Glasgow	1-Jul-12	31-Aug-13	350	CO40	Tina Murphy
CR76	PG16	John Kay	5 Novar Dr, Glasgow	1-Sep-13	1-Sep-14	450	CO93	Tony Shaw
CR56	PG4	Aline Stewart	6 Lawrence St, Glasgow	1-Sep-11	10-Jun-12	350	CO40	Tina Murphy
CR56	PG36	Aline Stewart	2 Manor Rd, Glasgow	10-Oct-12	1-Dec-13	375	CO93	Tony Shaw
CR56	PG16	Aline Stewart	5 Novar Dr, Glasgow	1-Nov-14	10-Aug-15	450	CO93	Tony Shaw

Figure 14.11 First Normal Form ClientRental relation.

# ФЗ за релацията ClientRental

#### ClientRental clientNo propertyNo cName pAddress rentStart rentFinish ownerNo oName rent fd1 (Primary key) (Partial dependency) fd2 (Partial fd3 dependency) (Transitive fd4 dependency) fd5 (Candidate key) fd6 (Candidate key)

Figure 14.12 Functional dependencies of the ClientRental relation.

# ФЗ за релацията ClientRental

```
 fd1 clientNo, propertyNo → rentStart, rentFinish
 fd2 clientNo → cName
 fd3 propertyNo → pAddress, rent, ownerNo, oName
 fd4 ownerNo → oName
 fd5 clientNo, rentStart → propertyNo, pAddress, rentFinish, rent, ownerNo, oName
 fd6 propertyNo, rentStart → clientNo, cName, rentFinish
```

# Ключ за релацията ClientRental

- Има три възможности за първичен ключ на релацията:
  - ▶ 1. (clientNo, propertyNo) PK
  - ▶ 2. (clientNo, rentStart) кандидат ключ
  - ▶ 3. (propertyNo, rentStart) кандидат ключ

#### Доказателство:

 Намираме покритието за трите двойки и се уверяваме, че това е така

### Нормализираме до НФБК

```
 fd1 clientNo, propertyNo → rentStart, rentFinish
 fd2 clientNo → cName
 fd3 propertyNo → pAddress, rent, ownerNo, oName
 fd4 ownerNo → oName
 fd5 clientNo, rentStart → propertyNo, pAddress, rentFinish, rent, ownerNo, oName
 fd6 propertyNo, rentStart → clientNo, cName, rentFinish
```

Функционалните зависимости fd2 нарушава правилото от НФБК. Декомпозираме по алгоритъма ClientRental и получаваме:

- Clients(clientNo, cName) ΗΦΕΚ
- ClientRental2(clientNo, propertyNo, pAddress, rent, ownerNo, oName, rentStart, rentFinish)

### Нормализираме до НФБК

```
 fd1 clientNo, propertyNo → rentStart, rentFinish
 fd2 clientNo → cName
 fd3 propertyNo → pAddress, rent, ownerNo, oName
 fd4 ownerNo → oName
 fd5 clientNo, rentStart → propertyNo, pAddress, rentFinish, rent, ownerNo, oName
 fd6 propertyNo, rentStart → clientNo, cName, rentFinish
```

Функционалните зависимости fd3 нарушава правилото от НФБК. Декомпозираме по алгоритъма ClientRental2 и получаваме:

- Properties(propertyNo, pAddress, rent, ownerNo, oName)
- ClientRental3(clientNo, propertyNo, rentStart, rentFinish) -НФБК

### Нормализираме до НФБК

```
 fd1 clientNo, propertyNo → rentStart, rentFinish
 fd2 clientNo → cName
 fd3 propertyNo → pAddress, rent, ownerNo, oName
 fd4 ownerNo → oName
 fd5 clientNo, rentStart → propertyNo, pAddress, rentFinish, rent, ownerNo, oName
 fd6 propertyNo, rentStart → clientNo, cName, rentFinish
```

Функционалните зависимости fd4 нарушава правилото от НФБК. Декомпозираме по алгоритъма Properties и получаваме:

- Properties(propertyNo, pAddress, rent, ownerNo) ΗΦΕΚ
- Owners(ownerNo, oName) ΗΦΕΚ

#### НФБК

#### Client

clientNo	cName
CR76	John Kay
CR56	Aline Stewart

#### Rental

clientNo	propertyNo	rentStart	rentFinish
CR76	PG4	1-Jul-12	31-Aug-13
CR76	PG16	1-Sep-13	1-Sep-14
CR56	PG4	1-Sep-11	10-Jun-12
CR56	PG36	10-Oct-12	1-Dec-13
CR56	PG16	1-Nov-14	10-Aug-15

#### PropertyForRent

propertyNo	pAddress	rent	ownerNo
PG4	6 Lawrence St, Glasgow	350	CO40
PG16	5 Novar Dr, Glasgow	450	CO93
PG36	2 Manor Rd, Glasgow	375	CO93

#### Owner

ownerNo	oName
CO40	Tina Murphy
CO93	Tony Shaw

JobID	JobDate Time	driverID	driver Name	taxiID	clientID	clientName	jobPickUpAddress
1	25/07/14 10.00	D1	Joe Bull	T1	C1	Anne Woo	1 Storrie Rd, Paisley
2	29/07/14 10.00	D1	Joe Bull	T1	C1	Anne Woo	1 Storrie Rd, Paisley
3	30/07/14 11.00	D2	Tom Win	T2	C1	Anne Woo	3 High Street, Paisley
4	2/08/14 13.00	D3	Jim Jones	T3	C2	Mark Tin	1A Lady Lane, Paisley
5	2/08/14 13.00	D4	Steven Win	T1	C3	John Seal	22 Red Road, Paisley
6	25/08/14 10.00	D2	Tom Win	T2	C4	Karen Bow	17 High Street, Paisley

- Нормализирайте до НФБК
- Решение:
  - ФЗ в сила за релацията са:
    - i. jobID->jobDateTime, jobPickUpAddress
    - 2. driverID -> driverName, taxiID
    - clientID -> clientName
  - ► Ключът на релацията е {jobID, driverID, clientID}
  - Всички Ф3, нарушават правилото на НФБК => декомпозираме

#### Нормализация

TaxiJob(jobID, jobDateTime, jobPickUpAddress, driverID, driverName, taxiID, clientID, clientName)

- Ф3-1, нарушава правилото за НФБК за ТахіЈоь, декомпозираме:
  - I. Jobs(jobID, jobDateTime, jobPickUpAddress) НФБК
  - 2. TaxiJob1(jobID, driverID, driverName, taxiID, clientID, clientName)
- Ф3-2, нарушава правилото на НФБК за ТахіJоb1, декомпозираме:
  - I. Drivers(<u>driverID</u>, driverName, taxiID) НФБК
  - 2. TaxiJob2(jobID, driverID, clientID, clientName)
- Ф3-3, нарушава правилото на НФБК за ТахіJob2, декомпозираме:
  - L Clients(<u>clientID</u>, clientName) НФБК
  - 2. TaxiJob3(<u>jobID</u>, <u>driverID</u>, <u>clientID</u>) НФБК

# Как да решаваме задачи за нормализация?

• Задача: Дадена е релацията R и описание към нея. Намерете всички Ф3, в сила за релацията. Нормализирайте до НФБК.

#### Алгоритъм:

- 1. Намирате всички Ф3, които са в сила за релацията. Извличате ги или от условието на задачата или от факти от действителността. Например ЕГН-то еднозначно определя всеки живущ в България.
- 2. Определяте кандидат ключовете за релацията R
- 3. Търсите Ф3, която нарушава правилото на БК (Ф3, която в лявата си част не съдържа ключа). Например FD I
- Декомпозирате R до две релации RI и R2, така че всички атрибути от FDI са в RI, а лявата част на FDI и всички останали атрибути са в R2.
- 5. Повтаряме стъпка 3) и 4) за релацията R2

Student_Id	Student_Address	Lecture	Teaching_Assisant
1234	Rämistrasse 72	Data Modelling and Databases	Bob
1280	Rennweg 19	Concepts of Concurrent Computation	Scott
1234	Rämistrasse 72	Visual Computing	Sarah
1299	Börsenstrasse 42	Concepts of Concurrent Computation	Benjamin
1356	Klusplatz 45	Concepts of Concurrent Computation	Benjamin

#### Известно е че:

- ▶ Един курс (Lecture) се води само от един Teaching\_Assistant
- Намерете всички ФЗ за така зададената релация
- Нормализирайте до НФБК
- Отговор:
  - Students(<u>Student\_Id</u>, Student\_Address)
  - Lectures(<u>Lecture</u>, Teaching\_Assistant)
  - Student\_Lectures(<u>Student\_Id</u>, <u>Lecture</u>)

leaseNo	bannerID	placeNo	fName	IName	startDate	finishDate	flatNo	flatAddress
10003	B017706	78	Jane	Watt	01/09/2010	30/06/2011	F56	34 High Street, Paisley
10259	B017706	88	Jane	Watt	01/09/2011	30/06/2012	F78	111 Storrie Road, Paisley
10364	B013399	89	Tom	Jones	01/09/2011	30/06/2012	F78	111 Storrie Road, Paisley
10566	B012124	102	Karen	Black	01/09/2011	30/06/2012	F79	120 Lady Lane, Paisley
11067	B034511	88	Steven	Smith	01/09/2012	30/06/2013	F78	111 Storrie Road, Paisley
11169	B013399	78	Tom	Jones	01/09/2012	30/06/2013	F56	34 High Street, Paisley

- Известно е че в релацията се съхранява информация за студенти (fName, IName) и в коя квартира са настанени в общежитието (flatNo, flatAddress) и на кое място (placeNo). Всеки студент еднозначно се определя от (bannerID)
- Нормализирайте до НФБК

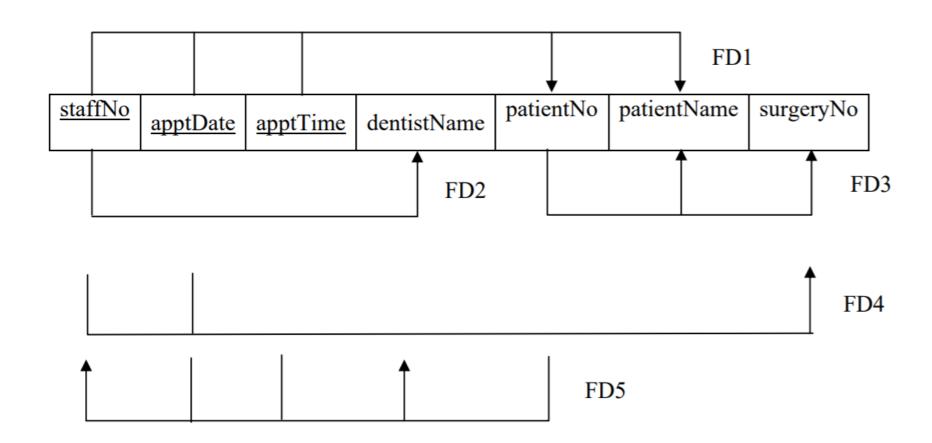
#### Задача 2 - Решение

- StudFlat(leaseNo, bannerID, placeNo, fName, IName, startDate, finishDate, flatNo, flatAddress)
- ▶ FD-I: bannerID -> fName, IName
- FD-2: flatNo -> flatAddress
- FD-3: placeNo -> flatNo
- ► FD-4: leaseNo, bannerID, placeNo -> startDate, finishDate
- {leaseNo, bannerID, placeNo}+
- ► I)X={leaseNo, bannerID, placeNo}
- 2)X={leaseNo, bannerID, placeNo, startDate, finishDate, fName, IName, flatNo, flatAddress}
- PK:{leaseNo, bannerID, placeNo}
- BCNF:
- Students(bannerID, fName, IName)
- Flats(flatNo, flatAddress)
- Places(placeNo, flatNo)
- Leases(leaseNo, bannerID, placeNo, startDate, finishDate)

 Дадена е релацията. Намерете всички Ф3, в сила за релацията. Нормализирайте до НФБК

staffNo	dentistName	patientNo	patientName	appointment	surgeryNo
				date time	
S1011	Tony Smith	P100	Gillian White	12-Aug-03 10.00	S10
S1011	Tony Smith	P105	Jill Bell	13-Aug-03 12.00	S15
S1024	Helen Pearson	P108	Ian MacKay	12-Sept-03 10.00	S10
S1024	Helen Pearson	P108	Ian MacKay	14-Sept-03 10.00	S10
S1032	Robin Plevin	P105	Jill Bell	14-Oct-03 16.30	S15
S1032	Robin Plevin	P110	John Walker	15-Oct-03 18.00	S13

# Задача 3 - ФЗ



### Задача 3 - НФБК

staffNo dentistName

Dentist(<u>staffNo</u>, dentistName)

FK

staffNo apptDate surgeryNo

Surgery(staffNo, apptDate, surgeryNo)

patientNo patientName

Patient(patientNo, patientName)

FK

staffNo apptDate apptTime patientNo

Appointment(staffNo, apptDate, apptTime, patientNo)

 Дадена е релацията. Намерете всички Ф3, в сила за релацията. Нормализирайте до НФБК

NIN	contractNo	hoursPerWeek	eName	hotelNo	hotelLocation
113567WD	C1024	16	John Smith	H25	Edinburgh
234111XA	C1024	24	Diane Hocine	H25	Edinburgh
712670YD	C1025	28	Sarah White	H4	Glasgow
113567WD	C1025	16	John Smith	H4	Glasgow

### Задача 4 - ФЗ

- ► FDI: NIN, contractNo -> hours
- ▶ FD2: NIN -> eName
- ▶ FD3: contractNo -> hotelNo, hotelLocation
- FD4: hotelNo -> hotelLocation

### Задача 4 - НФБК

- Staff(NIN, eName)
- ▶ Hotel(<u>hotelNo</u>, hotelLocation)
- Contract(<u>contractNo</u>, hotelNo)
- WorkHours(NIN, contractNo, hours)

 Дадена е релацията. Намерете всички Ф3, в сила за релацията. Нормализирайте до НФБК

StdSSN	StdCity	StdClass	OfferNo	OffTerm	OffYear	EnrGrade	CourseNo	CrsDesc
S1	SEATTLE	JUN	01	FALL	2006	3.5	C1	DB
S1	SEATTLE	JUN	O2	FALL	2006	3.3	C2	VB
S2	BOTHELL	JUN	O3	SPRING	2007	3.1	C3	00
S2	BOTHELL	JUN	O2	FALL	2006	3.4	C2	VB

#### Задача 5 - Решение

- StdClass(StdSSN, OfferNo, StdCity, StdClass, OffTerm, OffYear, CourseNo, CrsDesc, EnrGrade)
- FD-1: StdSSN -> StdCity, StdClass
- ▶ FD-2: OfferNo -> OffTerm, OffYear, CourseNo, CrsDesc
- FD-3: CourseNo -> CrsDesc
- FD-4: StdSSN, OfferNo -> EnrGrade
- {StdSSN, OfferNo}+ =
- {StdSSN, OfferNo, StdCity, StdClass, OffTerm, OffYear, CourseNo, CrsDesc, EnrGrade}
- PK: {StdSSN, OfferNo}
- BCNF:
- Students(StdSSN, StdCity, StdClass)
- Courses(CourseNo, CrsDesc)
- Offers(OfferNo, OffTerm, OffYear, CourseNo)
- Grades(StdSSN, OfferNo, EnrGrade)