Subject: اصول طراحی پایگاه داده ها

Year. 86 Month. 6 Date. 31 (1)

مهندس عبدالرضا پیرشایان

«به نام آن که آتش را آفرید تا سوز بی عشق آتش نگیرد.»

جلسه اول 86.6.31 ← فصل (1) →

DBMS : Data Base Management System

وظایف DBMS :

۱. حجم عظیمی از اطلاعات را در فایل نگه داری می کند.

۲. به وسیله یک سری از برنامه ها به این اطلاعات (داده های ذخیره شده) دسترسی پیدا می کند و این دسترسی به دو صورت انجام می گیرد :

- دسترسی DBMS به اطلاعات ذخیره شده
  - Convenient (راحت)
  - efficient (با صرفه / مقرون به صرفه)

معایب عدم وجود DBMS :

۱. Data redundancy (افزونگی داده ها) ← حافظه اضافی اشغال می شود و باعث حجم بالا می شود

۲. inconsistency (عدم همخوانی) ← سمت های مختلف سیستم با هم هماهنگی ندارد

۳. مشکل در دسترسی داده ها

E.F. Codd

Jim Gray قوانین جامعیت E.F.Codd را به 4 حالت زیر تبدیل کرد :

1) Atomicity (تجزیه ناپذیری) : یک عمل یا **کامل** انجام می شود و یا **اصلاً** انجام نمی شود.

2) Consistency (هم خوانی) : سمت های مختلف با هم هماهنگ هستند.

3) Isolation (جداپذیری) : همروندی کاربرها تاثیری (منفی) روی هم نداشته باشند (هر کاربر به تنهایی کارش را انجام دهد)

4) Durability (پایایی) : برای داده های ذخیره شده حفاظت امنیتی و فیزیکی صورت گیرد تا داده ها از بین نروند. (داده ها باید پایدار باشند)

PAPCO

Subject:
Year. Month. Date. ( )

Rbstraction : نگرش برداشت کلی از یک واقعیت را گویند یعنی به جزئیات توجهی نداریم

Data Abstraction : (سطوح داده ها انتزاعی)

1, Physical level ⇝ نحوه ذخیره شدن داده ها (داده چه جوری ذخیره می شود)
2, logical level ⇝ چه داده ای ذخیره می شود) نوع داده های ذخیره شونده
3, View level ⇝ براساس نیاز کاربر اطلاعات مورد نیاز را در اختیار می گذارد

level بالاترین

Schema : ساختار کلی data base بدون در نظر گرفتن داده ها را schema گویند
و کثیف ترین می آید که schema تغییر کند (یعنی رکوردها را حذف کرده باشیم)

Schema instance (نمونه) : نمونه هایی از schema در هر لحظه (یعنی داده ها در یک لحظه یک schema هستند)

Data Models:

Data Models مجموعه ای از ابزارهای مفهومی برای توصیف داده ها، ارتباط داده ها، مفاهیم و نقش داده ها (Semantic)، همخوانی و محدودیت (Constraint) داده ها برای حفظ یکپارچگی داده هاست؛

1, Relational Model
2, Entity - Relationship Model
3, Object - Based Model
4, Semi - Structured Model

جلسه دوم 86.7.7 ← فصل (6)

Entity Relationship Model: → این مدل بهترین ابزار برای طراحی Data Base است

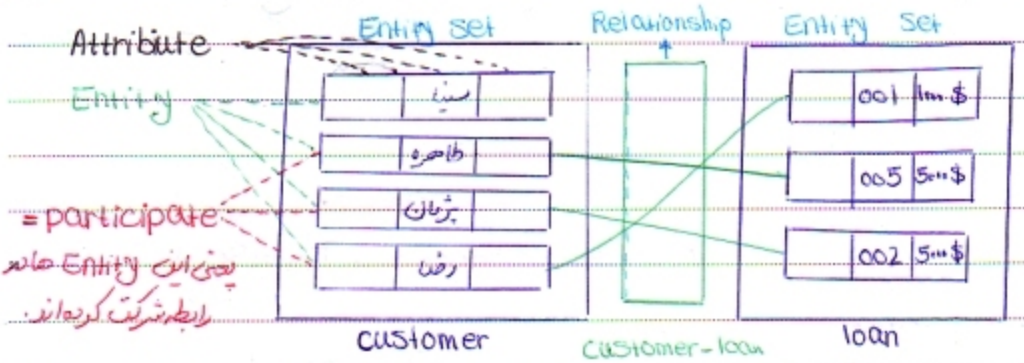
موجودیت ارتباط

- Entity: یک چیز، شی (object) است. مثلاً دانشجو، درس، استاد و... در سیستم دانشگاه Entity هستند.

- Set (مجموعه): گروهی از موجودیت ها (چیزها) که ویژگی مشترکی دارند را یک Set گوییم.

- Entity Sets: به مجموعه‌ای از Entity ها گفته می‌شود مثلاً مجموعه دروس، مجموعه اساتید و مجموعهٔ دانشجوها در سیستم دانشگاه Entity sets هستند.

مثال: در زیر دو Entity Sets بنامهای Customer (مشتری) و loan (وام) وجود دارد که هرکدام از آنها به ترتیب شامل 4 و 3 Entity هستند که بین این دو Entity Sets روابطی تعریف شده است این Relationship را Customer-loan می‌نامیم.



توجه: همانطور که می‌بینیم هر وام حتماً باید به کسی تعلق گیرد یعنی تمام وام ها participate هستند اما در مجموعهٔ مشتری‌ها می‌تواند شخصی باشد که وامی را دریافت نکرده (یعنی در رابطهٔ شرکت نکرده باشد) مثل سینا در این مثال.

Subject:
Year. Month. Date. ( )

Attributes: اطلاعاتی راجع به Entity به ما می‌دهد و به طور کلی به دسته‌های زیر تقسیم می‌شود:

- Simple (ساده): صفت ساده مثل اسم، شماره دانشجویی و...
- Composite (مرکب): صفتی که بتوان آن را به قسمت‌های کوچکتر تقسیم کرد مثل آدرس
- Single_valued (تک مقداری): صفتی که هر Entity تنها یک مقدار را بتواند قبول کند مثلاً شماره دانشجویی برای هر دانشجو یک صفت تک مقداری است.
- multi_valued (چندمقداری): صفتی که هر Entity بتواند چند مقدار آن را قبول کند مثل شماره تلفن که هر Entity می‌تواند بیش از یک شماره تلفن داشته باشد.
- Derived (مشتق شده): یعنی مقدار آن از .At دیگری گرفته شده باشد مثلاً .At سن که از روی .At تاریخ تولد محاسبه می‌شود (سن از تاریخ تولد مشتق شده) و یا معدل که از روی نمرات دروس مشتق شده است.

Relationship Set: در زیر به تعریف ریاضی مجموعهٔ روابط می‌پردازیم، باید توجه داشته باشیم که Relationship set مجموعه‌ای از چندتایی‌هاست.

Relationship Set:

$$\{(e_1, e_2, \dots, e_n) \mid e_1 \in E_1, e_2 \in E_2, \dots, e_n \in E_n\}$$

مجموعه‌ای — از چندتایی‌ها

entity: ✗ / Entity Sets ✓

- در مثال قبل دو Entity Sets داریم ← درجه 2 است.

⟸ تعداد Entity Set = درجه Relationship Set

مجموعه 2 تایی

customer_loan = { (طاهره ,005) (رضا ,001) (پژمان ,002) }

customer = { پژمان ، سینا ، رضا ، طاهره }

loan = {001, 002, 005}

* محدودیت‌ها (Constraints) عبارتند از:

PAPCO

Subject: اصول طراحی پایگاه داده ۱

Year. 86 Month. 7 Date. 7 13

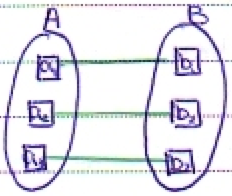
1* mapping Cardinalities → صحبت روی تعداد map ها را گویند

نگاشت   تعدادپذیری

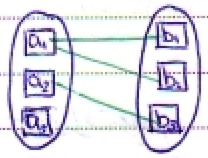
- mapping : نگاشت بین دو چیز و نسبت دادن را mapping نامیم.

- Cardinalites : یعنی تعدادپذیری، به این معنی که اجازه داشته باشیم یک چیز را چندبار map کنیم. مثلاً در مثال قبل به رضا هم وام 001 و هم وام 002 تعلق گیرد.

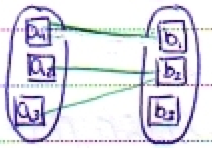
* در زیر به بررسی 4 mapping Cardinalities می پردازیم : (با توجه به مثال قبل)



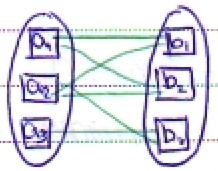
← One_to_One : هر مشتری تنها می تواند یک وام بگیرد و هر وام تنها می تواند به یک مشتری تعلق گیرد.



← One-to-many : هر مشتری می تواند صفر یا چند وام بگیرد ولی هر وام تنها باید به یک مشتری تعلق گیرد.



← Many to-One : هر مشتری می تواند حداکثر یک وام بگیرد و چند مشتری باهم می توانند یک وام بگیرند



← Many-to-many : هر مشتری می تواند به هر تعداد وام بگیرد و هر وام هم می تواند به هر تعداد مشتری تعلق گیرد (آزادترین حالت و بدون محدودیت)

2* participation Constraints: → (محدودیت های روابط)

شرکت کردن   محدودیت

مثلاً تعیین حداقل و حداکثر شرکت کردن هر مشتری در وام گرفتن و بدون در نظر گرفتن mapping)

P4PCO

Subject:
Year. Month. Date. ( )

توجه: باید دقت کنیم که participation Constraints و mapping Cardinalities کاملاً متفاوت است و نباید با هم اشتباه گرفته شوند.

*3 keys (کلیدها): کلید یک Entity Sets می تواند یک یا چند Attribute باشد که آن Attribute باید منحصر به فرد باشد.

- Super keys (ابرکلید): به کلیدی گفته می شود که بتواند به کلیدهای دلبسته شود یعنی چند کلید در یک مجموعه خود داشته باشد به طوری که به کلیدی super keys گفته می شود که در دل خود یک یا چند کلید داشته باشد. مثلاً اگر فرض کنیم نام و نام خانوادگی و شماره دانشجویی کلید باشد حتماً super keys خواهد بود.

- Candidate keys: به کلیدی که قابلیت انتخاب شدن دارد را candidate keys گوییم، درواقع super keys که شکسته نشود را candidate keys نامند. مثلاً کلید (اسم، نام خانوادگی، شماره دانشجویی) یک candidate keys است زیرا چنانچه این کلید را بشکنیم و هر قسمتش را برداریم دیگر کلیدی نخواهیم داشت.

توجه: بهتر است Attribute کمتری را به عنوان candidate keys در نظر بگیریم، مثلاً بین (نام و نام خانوادگی و شماره شناسنامه) و (شماره دانشجویی) بهتر است شماره دانشجویی را انتخاب کنیم.

- primary key: انتخاب یک key از بین چند candidate keys یک primary key است؛ در مثال بالا شماره دانشجویی یک primary key است.

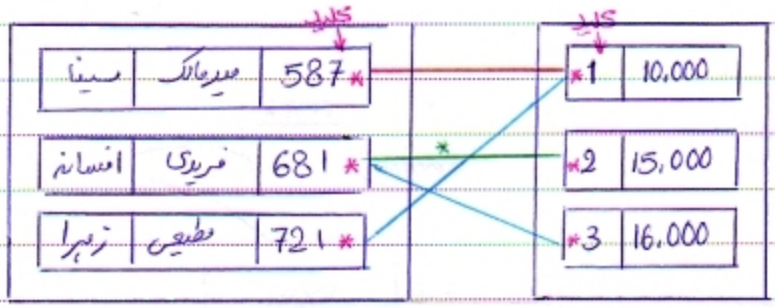
جلسه سوم 86.7.14

نکته: primary key ⊆ candidate key ⊆ super key

- هر super key ممکن است به کلیدهای دیگر تجزیه شود و باید توجه داشته باشیم که هر primary key یک super key هم هست.

نکته: در Relationship set برای ما primary key محسوب است

مثال: دو Entity set زیر را در نظر بگیرید



برای نوشتن مجموعه Relationship باید primary key را در نظر بگیریم.

Relationship set: {(681, 2), (721, 1)}

با توجه به mapping cardinalits عبارت است زوج (681, □) تکرار شود.

one-to-one: هیچ زوجی تکرار نمی‌شود و کلید ارتباط می‌تواند کلید Customer یا loan باشد

* one-to-many: زوج تکراری خواهد داشت؛ مثل (681, 2) (681, 3)

* many-to-one: زوج تکراری خواهد داشت؛ مثل (721, 1) (587, 1)

- همانطور که می‌بینیم در سمتی که one است تکرار وجود دارد بنابراین برای حل این مشکل کلید سمت many را به عنوان کلید ارتباط در نظر می‌گیریم.

و در حالت many-to-many اجتماع کلید هر دو سمت کلید ارتباط را تشکیل می‌دهند.

Entity Relationship Diagrams:

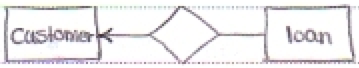
- قراردادهای رسم Diagram به صورت زیر است:
- مستطیل: Entity set را نشان می‌دهد.
- لوزی: Relationship را نشان می‌دهد.
- بیضی: Attribute را نشان می‌دهد.

خطوط: قسمت‌های مختلف diagram را به هم وصل می‌کنند:

—— : نشان دهندهٔ many است.

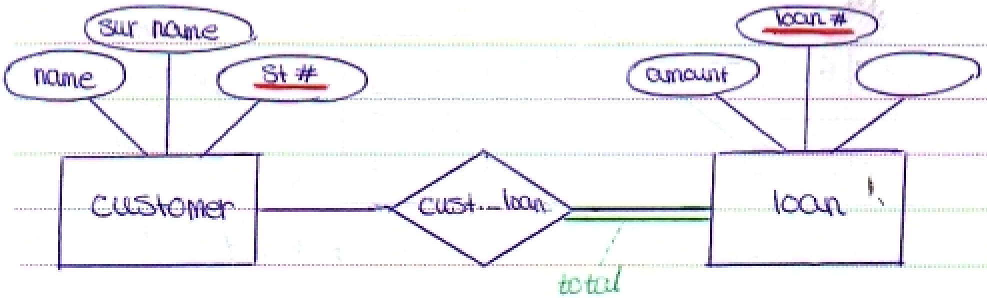
→ : نشان دهندهٔ one است.

نوک فلش به سمت مستطیل



one-to-many : از customer به loan

many-to-one : از loan به customer



total: در Entity set که همهٔ Entity هایش باید حتماً (حداقل) یکبار در رابطه شرکت کنند آنگاه آن Entity set را total می‌نامند و خط دوبل نشان داده می‌شود.

مثلاً در اینجا هر وامی باید حداقل یکبار در ارتباط شرکت کند.

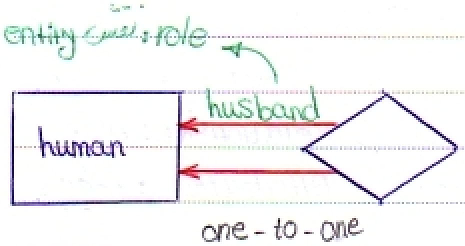
Double ellipses: multivalued att. را نشان می‌دهد.

Dashed ellipses: derived att. را نشان می‌دهد.

l..h : تعیین حداقل و حداکثر شرکت کردن را نشان می‌دهد.

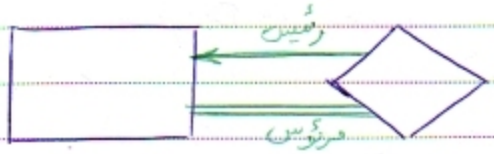
مثلاً 0..5 حداقل 0 و حداکثر 5 بار می‌تواند شرکت کند.

و 1..1 یعنی فقط و فقط یکبار می‌تواند در ارتباط شرکت کند.



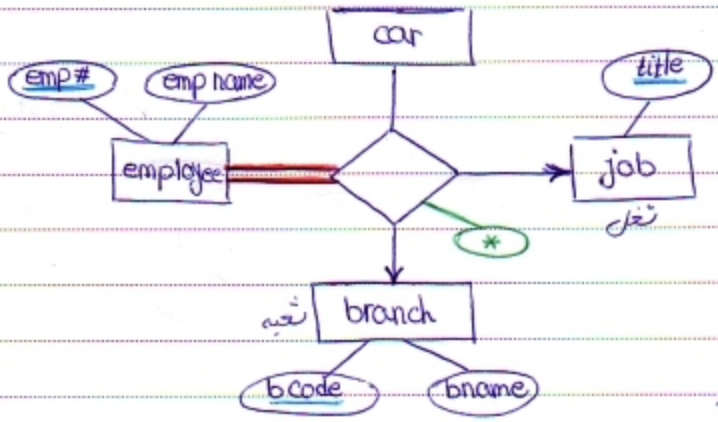
نکته: در برخی موارد لازم است از روابط بازگشتی استفاده کنیم. مثلاً اگر مجموعهٔ انسان‌ها را در نظر بگیریم برای نمایش ارتباط زناشویی، ارتباط دوستی و ... باید از خطوط بازگشتی مطابق شکل استفاده کنیم زیرا زن و شوهر هر دو در همان مجموعه انسان‌ها هستند.

مثال: رابطه رئیس و مرئوس بدون هیچ یک ارتباط بازگشتی است با توجه به اینکه کسی نمی‌تواند در ارتباط مرئوس بدون شرکت نکند بنابراین total است. و فرض می‌کنیم هر کس یک رئیس مستقیم داشته باشد.



many-to-one از مرئوس به رئیس
one-to-many از رئیس به مرئوس

مثال:
در اینجا یک Relationship چهارتایی داریم



* همانطور که می‌بینیم می‌توانیم برای Relationship ها هم attribute در نظر بگیریم و این پدیده وقتی رخ می‌دهد که آن Relationship به هیچ entity set مربوط نباشد.

* total است زیرا هر کارمند باید job داشته باشد یعنی در همه ارتباطات باید شرکت کند زیرا در غیر این صورت دیگر شخص کارمند نیست.

توجه: این مثال دارای ابهام است زیرا دارای دو فلش است (one) مثلاً می‌توان ارتباط سه‌تایی ترکیب از employee و car ــــ با یک ترکیب از job و branch تعریف کرد. یا یک job و یک branch تعریف کرد.

برای رفع ابهام در ارتباطات بیش از دوتایی باید تنها یک ارتباط را با فلش تعریف کنیم. در غیر این صورت بهم خواهد بود.

Subject :
Year . Month . Date . ( )

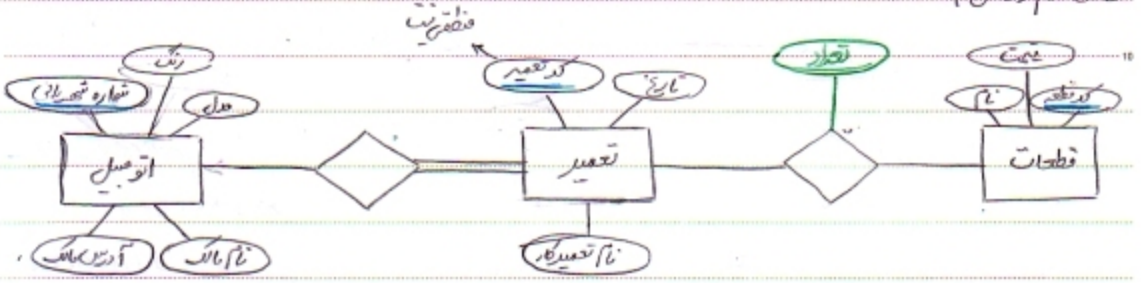
مثال : Diagram تعمیرگاهی را رسم کنید که اطلاعات مشتریها، اتومبیل ها و قطعات تعویضی در آن نشان داده شود.

- می توانیم مشتری را یک entity set جدا کنیم و یا مشتری را به عنوان یک attribute اتومبیل در نظر بگیریم که حالت دوم بهتر است.

حالت اول (نامناسب)

هر ماشین متعلق به یک نفر است ولی هر مشتری می تواند چند ماشین داشته باشد.

حالت دوم (کامل)



تعداد قطعات هم به تعمیر و هم به قطعات بستگی دارد بنابراین روی Relationship قرار می گیرد.

- توجه : اگر ارتباط اولی را به این دنیا می گرفتیم و قطعات را هم به آن متصل می کردیم Diagram اشتباه بود زیرا در آن صورت هر ماشینی که به تعمیرگاه می رفت باید قطعه ای را هم تعویض می کرد در صورتی که امکان دارد اتومبیلی بدون تعویض قطعه تعمیر شود.
- تعداد را نمی توانیم روی قطعات تعریف کنیم زیرا در آن صورت مشخص نمی شود که کدام قطعه مربوط به کدام تعمیر (ماشین) است.

نکته : می توان Entity Relationship Diagram را به table متصل کرد (به راحتی) به همین دلیل ER بسیار مورد توجه و استفاده قرار گرفته است.

مثال: می خواهیم مثال قبل را به صورت جدول تبدیل کنیم.

در تبدیل Diagram به table:

هر Entity set به یک جدول و هر Relationship set نیز یک جدول انتقال می یابد. ستونهای جدول Entity set attribute های Entity set را تشکیل می دهند و ستونهای جدول Relationship set را کلید ارتباط و همچنین at. های روی ارتباط تشکیل می دهند و سطرهای این جداول شرکت کننده ها هستند.

جدول تعمیر

نام تعمیرکار	تاریخ	کد تعمیر
رضا	۸۹/۷/۱۴	1
بهمن	۸۹/۹/۹	2
حسن	۸۹/۳/۵	3
محسن	۸۹/۷/۹	4

جدول اتومبیل

آدرس مالک	نام مالک	مدل	رنگ	شماره شهربانی
تهران	علی	۱۳۷۹	سفید	۲۵ م ۹۴۵
تهران	سارا	۱۳۸۹	نقره‌ای	۱۵ ق ۱۴۱
تهران	مینا	۱۳۸۳	سبز	۴۵ م ۸۹۹

جدول قطعات

کد قطعه	نام	قیمت
1	شمع	
2		
3		

- چون رابطه اول many-to-many کلید هردو Entity set را باید در جدول بنویسیم ← (دو ستون)
- در ارتباط دوم (سه ستون) داریم زیرا علاوه بر دو کلید یک ستون هم به at. تعداد تعلق می گیرد.

کلید اتومبیل ← کلید تعمیر ← کلید قطعات ← at

شماره شهربانی	کد تعمیر	کد قطعه	تعداد
۲۵م ۶۴۵	2	1	1
۲۵ م ۶۴۵	3	2	4
۴۵ م ۸۹۹	1		
۱۵ ق ۱۴۱	4		

جدول ارتباط (R) اول

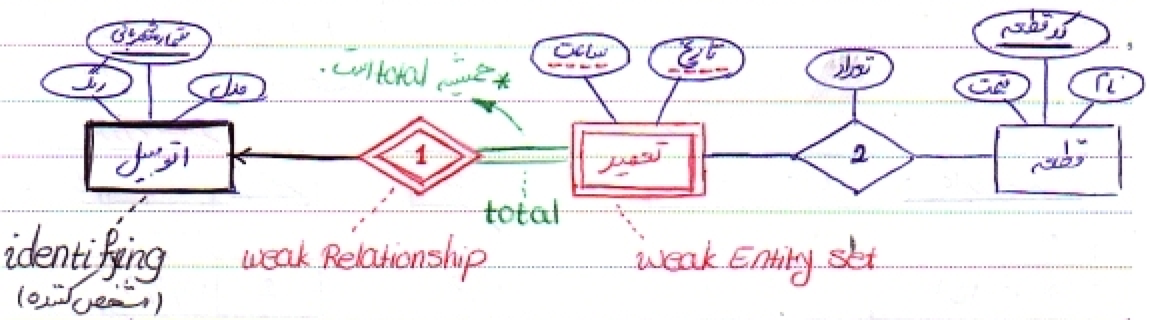
جدول ارتباط (R) دوم

Subject:
Year. Month. Date.

جلسه چهارم 86.7.28

Weak Entity Sets : entity set که استقلال ندارد و به entity set های دیگر وابسته است

انتیتی ضعیف گفته می‌شود چرا که هویت وابسته دارد. (primary key ندارد)

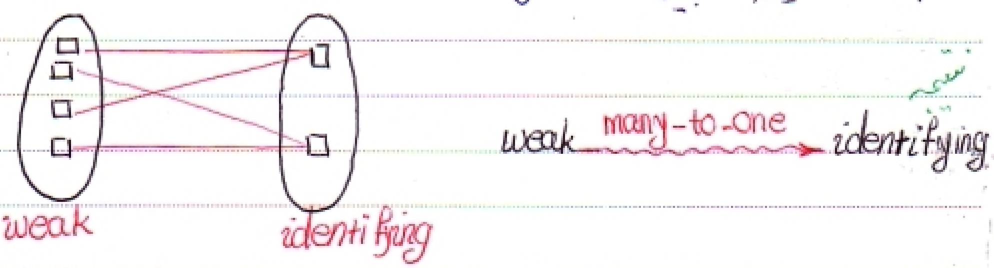


- کلید weak entity set را partial key می‌نامند چون واقعاً کلید نیست اما در تشکیل کلید دخالت دارد مثلاً در اینجا (ساعت، تاریخ) partial key هستند.

توجه: برای تعیین کلید weak entity set کافی است partial key + کلید attribute وابسته را باهم در نظر بگیریم، یعنی در اینجا کلید انتیتی ست تعمیر (ساعت، تاریخ، شماره شهربانی) است.

نکته: اگر چند attribute باهم یک primary key را در یک Entity set تشکیل دهند آن Entity set نمی‌تواند weak باشد مثلاً چنانچه در مثال قبلی attribute دیگری به عنوان نام تعمیرکار هم بود دیگر تعمیر یک weak ent. نبود.

* هر Entity در Entity Set تعمیر فقط و فقط به یک اتومبیل وابسته است بنابراین ارتباط از تعمیر به اتومبیل many-to-one است.



PAPCO

Strong Entity Set : هر انتیتی ستی که weak نباشد Strong است یعنی انتیتی ستی که primary key دارد و مستقل است Strong گفته می‌شود.

- مثال قبل را در نظر بگیرید:

جدول تعمیر

شماره شهربانی	تاریخ	ساعت

جدول ارتباط (۱)

شماره شهربانی	تاریخ	ساعت

قابل حذف

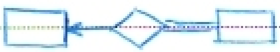
جدول ارتباط (۲)

شماره شهربانی	تاریخ	ساعت	کد قطعه	تعداد

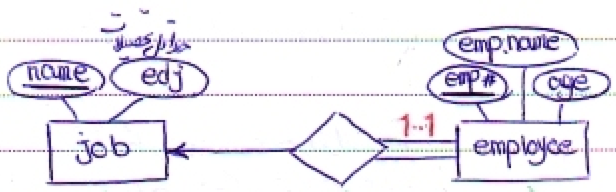
- وقتی می‌توانیم جدولی را حذف کنیم که اضافی باشد یعنی از حذف اطلاعاتی از سیستم پاک نشده خدشه‌ای به سیستم وارد نکند در اینجا هم همانطور که می‌بینیم Table ارتباط (۱) و تعمیر کاملاً مثل هم بوده یکی از آنها قابل حذف است؛

دلیل شباهت دو جدول: ۱- ارتباط مربوط به Entity Set تعمیر total است و ۲- ارتباط many-to-on از تعمیر به اتومبیل؛ سبب شده این دو جدول کاملاً شبیه باشند.

← هرجا ارتباط many-to-one داشته باشیم که many اش total باشد می‌توانیم جدول Relationship را حذف کنیم.



به مثال زیر توجه کنید:



name	edj

name	emp#

قابل حذف

emp#	age	emp.name

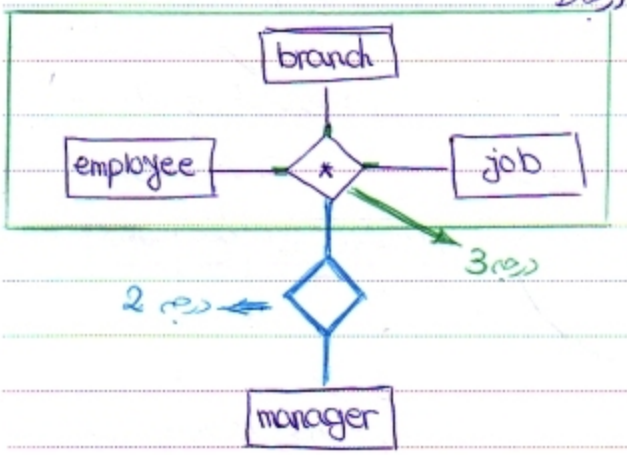
name	emp#	age	emp.name

طبق نکته ذکر شده در بالا جدول ارتباط قابل حذف است اما باید ستون name را به جدول employee اضافه کنیم.

Subject :
Year . Month . Date . ( )

: (Extended ER) EER

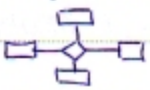
aggregation : (تجمع) یعنی یک نسبت ارتباط



* این ارتباط مشخص می کند که چه کسی ، چه شغلی دارد و در کدام شعبه مشغول به کار است ؛ ممکن است یک شخص در دو شعبه در دو شغل متفاوت داشته باشد اما در هر شعبه یک رئیس جدا گانه دارد بنابراین نمی توانیم بین کارمند و رئیس relationship بگذاریم لذا مجبوریم از طریق تعریف یک Relationship دیگر Entity رئوسا را به این Diagram اضافه کنیم .

((شعبه ، شغل و کارمند) و رئیس)

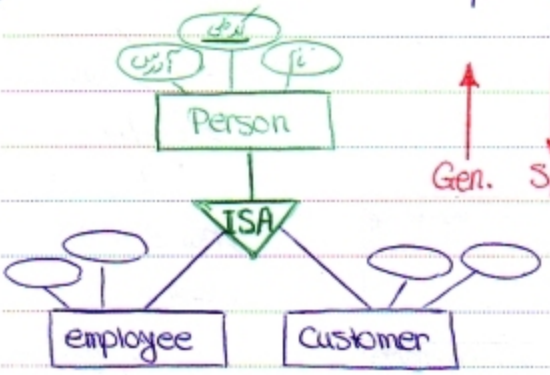
و در این ارتباط تعریف نکنیم ارتباط * در 4 خواهد بود و در آن صورت هر ارتباط با 4 icon مشخص می شود (شعبه ، شغل ، کارمند ، رئیس) که اشتباه است .



: Specialization & Generalization

جزئی ، کلی

گاهی اتفاق می افتد که دو Entity Set دارای attribute های مشترک و کلیدی باشند در این صورت بهتر است یک انتیتی ست جدا تعریف کنیم و att. های مشترک را به آن اختصاص دهیم .



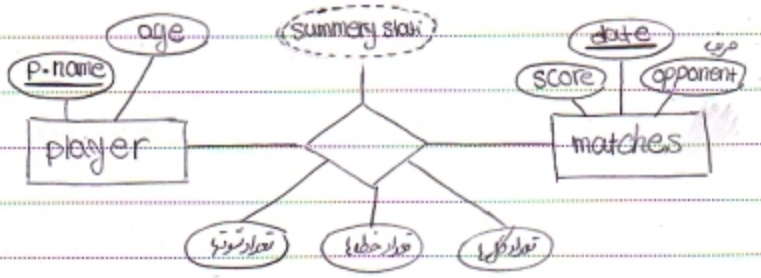
مثال : مثلاً کارمند و مشتری هر دو انسانند و دارای att. مشترکی مثل نام ، آدرس ... هستند بنابراین مانند شکل رو برد آن را پیاده سازی می کنیم :

یعنی از  استفاده می کنیم .

Subject: اصول طراحی پایگاه داده ۴
Year. 86 Month. 7 Date. 28 8

تمرین P. 256 / 6.4

دیاگرامی رسم کنید که تیم مورد علاقه تان ، سابقاتش ، اسامی که در هر سابقه کسب می کند ، بازیکنان تیم در هر سابقه و آمار بازی ذخیره شود.



- داریم دنبال یک تیم (تیم مورد علاقه مان) هستیم بنابراین لازم نیست مثلاً اسم تیم را ذخیره کنیم فقط کافی است نام حریف را ذخیره کنیم و دراین صورت date می تواند کلید باشد چون یک تیم در یک تاریخ فقط می تواند دیدار مسابقه ای داشته باشد بنابراین date برای matches یک primary key است.

و آن جایی که att. هایی مثل تعداد خطاها، تعداد توپ ها ... تیم به بازیکن و تیم به سابقه بستگی دارد باید آنها را دری Relationship تعریف کرد.

جلسه پنجم 86. 8. 5 ← فصل (2)

Relational Model : این مدل، یک مدل موفق است (این مدل برای درخواستهای تجاری) و مهم ترین عامل در این مدل ارتباطات است.

Domain : دامنهٔ هر attribute مقادیر مجازی است که هر att. می تواند داشته باشد (یعنی مقادیر مجازی که در هر ستون می تواند قرار گیرد)

header (هر ستون نام attribute است)

توجه: هر relational DB مجموعه ای از جداول است که هر جدول یک نام مشخص دارد و هر ردیف (row) این جدول یک relationship را نشان می دهد و به طور غیر مستقیم هر جدول یک Entity set و هر ستون یک attribut است.

Subject:
Year. Month. Date. ( )

Basic Structure : در این جدول account relation را می بینیم که سه ستون دارد که
D_1 : مقادیر مجاز account # ، D_2 : مقادیر مجاز branch-name و
D_3 : مقادیر مجاز balance را شامل می شود و
هر سطر جدول شامل سه tuple (v_1, v_2, v_3) است. tuple = ردیف های جدول

← بنابراین account زیرمجموعه ای از $D_1 \times D_2 \times D_3$ است.

نتیجه : در حالت کلی، یک table شامل n attribute زیرمجموعه ای از
$D_1 \times D_2 \times \dots \times D_n$ می باشد.

account

account #	branc-name	balance
A_1	میرداماد	۱۰۰۰۰۰۰
A_2	ونک	∅
R_3	ونک	۱۰۰۰۰

رابطه = $D_1 \times D_2 \times D_3$
رابطه = {A_1, A_2, A_3} × {ونک، میرداماد}
× {۱۰۰۰۰۰۰ ، ∅ ، ۱۰۰۰۰}
= {(A1، میرداماد، ۱۰۰۰۰۰۰) (A2، ونک، ∅) (A3، ونک، ۱۰۰۰۰)}
tuple tuple

نکته : برای نمایش tuple variable از نماد t[ ] استفاده می شود.

t [account #] = "A_1"
t [branch-name] = " میرداماد "

مثال : tuple varible relation
t ∈ r

Schema : نشان می دهد رابطه چه att. هایی دارد و با حرف بزرگ آغاز می شود ؛ مثال،
Account_schema = (account #, branch-name, balance)
int مثل
a مثل
account (Account_schema)
account (account #, branch-name, balance) OR
نام جدول کلید
در اینجا یک رابطه از نوعی که خودمان تعریف کردیم را ایجاد کردیم مثل int a,b
relation

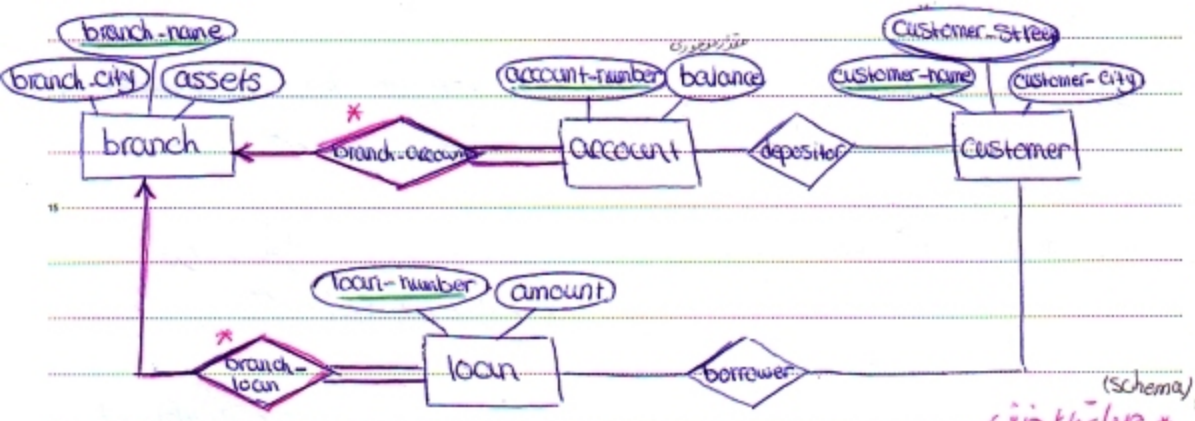
یادآوری: همانطوریکه قبلاً اشاره کردیم هر tuple variable با نماد t[ ] نشان داده می‌شود (یک سطر از یک جدول که ممکن است در هر لحظه یک مقدار داشته باشد)

R_Stance : مقادیری که در ارتباط مورد نظرمان در هر لحظه می‌گیرد.

اگر R یک Relational schema و k یک super key باشد داریم:

$$t_1 \neq t_2 \Rightarrow t_1[k] \neq t_2[k] \qquad K \subseteq R$$

مثال:



(Schema)

* جدول با کلید خارجی

branch (branch-name, branch-city, assets)

account (account-number, balance, branch-name)

customer (customer-name, customer-street, customer-city)

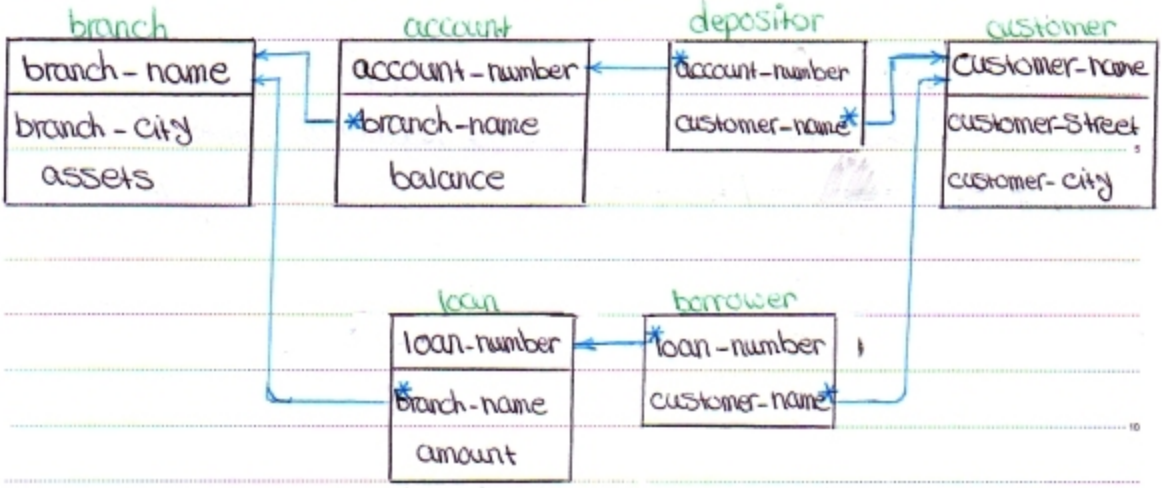
loan (loan-number, amount, branch-name)

depositor (customer-name, account-number)

borrower (customer-name, loan-number)

Subject:
Year. Month. Date. ( )

حال به Schema Diagram مثال قبل توجه کنید:



* این مقادیر foreign key هستند یعنی مقادیرشان وابسته بوده و از جداول دیگر تامین می‌گردد. یعنی در یک schema دیگر کلید هستند.

توجه: در این دیاگرام زیر کلیدها خط می‌کشیم و از آنجایی که depositor, borrower دو کلید دارند هر دو آنها را به عنوان کلید زیرشان خط می‌کشیم.

بازیابی اطلاعات:

به طور کلی سه نوع پرس‌وجو در مدل رابطه‌ای وجود دارد:

1. Relational Algebra (جبر رابطه‌ای)
2. Tuple Relational calculats (حساب رابطه‌ای رکوردی (تاپلی)) TRC
3. Domain Relational calculats (حساب رابطه‌ای دامنه‌ای) DRC

1. Relational Algebra: جبر رابطه‌ای دارای 6 عملگر اصلی (operations) می‌باشد که سه‌تای آنها unary و سه‌تای دیگر binary هستند عملگرهای

P4PCO

عملوند = (رابطه)
روی یک relation اجرا می‌شوند و عمگرهای binary روی دو عمگر اجرا می‌گیرند؛

و این شش عمگر عبارتند از: شرط

1. Selection (unary) → $\sigma_p(r)$ روی Relation
2. projection (unary) → $\Pi_A(r)$ $A \subseteq R$
3. Union (binary) → $\cup$
4. Set difference (binary) → $-$
5. Cartesiomproduct (binary) → $\times$
6. rename (unary → $\rho_x(E)$ exp — نام دلخواه

1، Selection: این عمگر unary به ما این امکان را می‌دهد که tuple هایی با شرایط خاص (دلخواه‌مان) را انتخاب کرده و آنها را از بقیه جدا کنیم؛ در قسمت شرط σ می‌توانیم از علائم $=, \neq, >, <, \geq, \leq$ ، $\wedge$ (and)، $\vee$ (or)، $\neg$ (not) استفاده کنیم.

مثلاً می‌خواهیم کسانی که در شعبهٔ vanak بوده و مقدار وامشان از 1000 بیشتر است را انتخاب کنیم

* $\sigma_{amount > 1000 \wedge branch\text{-}name = "vanak"}(loan)$ *

که جدول به صورت زیر خواهد شد:

loan-number	branch-name	amount
L-15	vanak	1500
L-16	vanak	1300

2، projection: این عمگر unary این امکان را به ما می‌دهد که ستونی که به آن احتیاج نداریم را از جدول حذف کنیم مثلاً فرض کنید می‌خواهیم لیست همهٔ شماره وام‌ها را تهیه کنیم و شعبه برای ما اهمیت ندارد در آن صورت داریم:

* schema تعیین می‌کند: $\Pi_{loan\text{-}number, amount}(loan)$ ⇒

و با اعمال این دستور در واقع ستون مربوط به branch-name حذف شده و جدول موردنظر ما با 2 ستون ایجاد خواهد شد.

Subject:
Year. Month. Date. ( )

مثال: می خواهیم نام مشتریهایی که در "Harrison" زندگی می کنند را بیابیم.
برای این منظور باید از ترکیب دو عملگر به صورت زیر استفاده کنیم:

$$\Pi_{customer\text{-}name}\left(\sigma_{customer\text{-}city = \text{"Harrison"}}(customer)\right)$$

انتخاب مشتریهایی که در Harrison زندگی می کنند

جداکردن ستون نام مشتری ها و نمایش آن

همانطور که می بینیم باید ستون نام مشتریها را توسط Π از بقیه جدا کرده و با σ تمهیدات لازم را انتخاب کنیم تا در پایان یک جدول تک ستونی (نام مشتری) داشته باشیم.

3. اپراتور union: این عملگر (اجتماع) به ما این امکان را می دهد که چندگزاره مختلف (با شرایط گوناگون) را به طور همزمان در نظر بگیریم. مثلاً می خواهیم نام مشتریهایی که وام گرفته اند و یا نام مشتریهایی که حساب دارند را تعیین کنیم:

$$\Pi_{customer\text{-}name}(borrower) \cup \Pi_{customer\text{-}name}(depositor)$$

(* borrower: کسانی که وام گرفته اند — * depositor: کسانی که حساب دارند)

* نکته: union بین مجموعه هایی استفاده می شود که schema های آنها از نظر تعداد و دامنه یکسان باشند

4. اپراتور set difference: این عملگر (تفریق) به ما اجازه می دهد tuple هایی را پیدا کنیم که در یک relation هستند اما در دیگری وجود ندارند. مثلاً می خواهیم نام مشتریهایی که حساب دارند ولی وام نگرفته اند را پیدا کنیم.

$$\Pi_{customer\text{-}name}(depositor) - \Pi_{customer\text{-}name}(borrower)$$

(* depositor: کسانی که حساب دارند — * borrower: کسانی که وام گرفته اند)

جلسه ششم 86.8.12

5. Cartesian-product : این عمگر به ما این امکان را می‌دهد که دو مجموعه رابطه را در هم ضرب (دکارتی) کنیم (۲ یا بیشتر) که نتیجه این ضرب دکارتی باید یک **رابطه** باشد.

مثال: فرض کنید می‌خواهیم دو رابطه loan و borrower را در هم ضرب کنیم:

borrower (loan-number, customer-name)

loan (loan-number, branch-nam, amount)

از آنجایی که loan-number در هر دو رابطه تکرار شده است باید در ضرب رفع ابهام کنیم:

borrower x loan =

(borrower.loan-number , borrower. customer-name ,

loan. loan-number , loan. branch-name, loan. amount)

چون بقیه موارد att. تکراری وجود ندارد بنابراین ابهامی نخواهیم داشت لذا دیگر احتیاجی نیست نام رابطه را در کنار att.ها بیاوریم.

نکته: در واقع ضرب دکارتی به ما این امکان را می‌دهد که ستونهای روابط مختلف را در کنار هم ایجاد کرده و در حالت‌های تجارت آنها را با هم مقایسه کنیم. هر چند ضرب یک عمل بسیار زمانبر و پرهزینه است.

borrower

loan-#	customer-name
L_1	آرین
L_2	علی
L_3	افسانه

loan

loan-#	branch-name	amount
L_1	میرداماد	10
L_2	ونک	20
L_3	میرداماد	30

borrower x loan

borrower. loan #	customer -name	loan. loan#	branch- name	amount
L_1	آرین	L_1	میرداماد	10
L_1	آرین	L_2	ونک	20
L_1	آرین	L_3	میرداماد	30
L_2	علی	L_1	میرداماد	10
L_2	علی	L_2	ونک	20
L_2	علی	L_3	میرداماد	30
L_3	افسانه	L_1	میرداماد	10
L_3	افسانه	L_2	ونک	20
L_3	افسانه	L_3	میرداماد	30

Subject:
Year. Month. Date. ( )

مثال: نام افرادی که از شعبه میرداماد وام گرفته اند را بیابید.

نام افرادی که وام گرفته اند در رابطه borrower و شعبه ای که وام گرفته اند در loan نگهداری می شود
بنابراین باید از ضرب دکارتی این دو مجموعه استفاده کنیم

در borrower x loan در جدول باید سطرهایی را انتخاب کنیم که loan# یکسان دارند
زیرا اگر در غیر اینصورت borrower.loan# و loan.loan# با هم یکی نباشند آن سطر بی معنا خواهد بود بنابراین داریم:

$$\Pi_{\text{customer-name}}\left(\sigma_{\substack{\text{branch-name} = \text{"میرداماد"} \\ \text{borrower.loan\#} = \text{loan.loan\#}}}\left(\sigma\left(\text{borrower} \times \text{loan}\right)\right)\right)$$

customer-name
آرین
امانه

6. Rename: این عملگر اصلی به ما این امکان را می دهد که نام هر exp صورت نظر حتی att های آن را در صورت نیاز تغییر نام دهیم. مثلاً فرض کنید می خواهیم borrower را در حودش ضرب کنیم در این صورت باید حتماً از عملگر rename استفاده کنیم زیرا در غیر این صورت اسامی تکراری خواهیم داشت.

$$\text{borrower} \times \text{borrower} = \text{borrower} \times \rho_{x}(\text{borrower})$$
$$= (\text{borrower.loan\#}, \text{borrower.customer-name}, x.\text{loan\#}, x.\text{cus}_{\text{name}})$$

توجه: می توانیم نام یک یا تک اجزای هر schema را نیز تغییر دهیم: مثال

$$\rho_{x(A_1, A_2)}(\text{borrower}) \Rightarrow x = \text{borrower};\quad A_1 = \text{loan-\#};\quad A_2 = \text{customer-name}$$

و چنانچه بخواهیم فقط نام یکی از att ها را تغییر دهیم باید نام بقیه را عیناً نوشته و نام مورد نظر را برای آن یک att بنویسیم. (مثلاً به جای A_1 همان loan# را بگذاریم)

تعریف رسمی از جبر رابطه ای: یک exp در جبر رابطه ای شامل ۱- رابطه در DB و ۲- یک رابطه ثابت (Constant) می باشد.

۱- اگر E_1 و E_2 در جبر رابطه ای expression باشند آنگاه تمامی روابط زیر نیز exp های جبر رابطه ای خواهند بود:

→ $E_1 \cup E_2$ → $E_1 - E_2$ → $E_1 \times E_2$

→ $\sigma_p(E_1)$ → $\Pi_A(E_1)$ → $\rho_x(E_1)$

عملگرهای اضافی در جبر رابطه‌ای:

علاوه بر 6 عملگر اصلی در جبر رابطه‌ای، 4 عملگر اضافی نیز در جبر رابطه‌ای وجود دارد که عبارتند از:

1. The set-Intersection Opration → "∩"
2. The Natural-join Opration → "⋈"
3. The Division Opration → "÷"
4. The Assignment Opration → "←"

1. Set_Intersection: اولین عملگر اضافی ∩ (اشتراک) است. فرض کنید می‌خواهیم مشتریهایی را پیدا کنیم که هم حساب دارند و هم وام گرفته‌اند، در این صورت داریم:

$$\Pi_{\text{customer-name}}(\text{borrower}) \cap \Pi_{\text{customer\_name}}(\text{depositor})$$

نکته: عملگر فرعی ∩ را می‌توان به کمک چند عملگر اصلی پیاده‌سازی کرد:

$$r \cap s = r - (r - s)$$

2. Natural join: در خروجی، att. های غیرمشترک و att. های مشترک (بدون تکرار) ظاهر خواهند شد؛ یعنی از att. های مشترک فقط یک نمونه نگه داری می‌شود.

این عملگر فرعی نیز ترکیبی از σ و × است:

$$r \bowtie s = \sigma_{r.A=s.A \wedge r.B=s.B \wedge \cdots}(r \times s)$$

شرط: att. های هم‌نام مقایسه‌شان سازگار باشد

$R \cap S = \{A, B, C, \ldots\}$ $r(R)$; $s(S)$;

مثال: نام مشتریهایی که از میرداماد وام گرفته‌اند؟

(فقط loan# مشترک است ← سه سطر انتخاب می‌شوند)

$$\Pi_{\text{customer-name}}\left(\sigma_{\text{branch-name} = \text{"میرداماد"}}(\text{borrower} \bowtie \text{loan})\right)$$

(از سه سطر، یک سطر خروجی می‌دهد)

Subject:
Year. Month. Date. ( )

نکته: علاوه بر Natural join یک عملگر دیگر به نام θ join نیز وجود دارد که تنها فرق این دو join در شرطشان است در θ join هر شرط دلخواهی می توانیم استفاده کنیم اما در Natural join شرط این است که att های همنامی که مقدارشان مساوی است مد نظر قرار گیرند.

$$\sigma_{\theta}(r \times s) = r \bowtie_{\theta} s$$

θ → شرط دلخواه

مثال: شماره حساب مشتری که max مقدار موجودی را دارد بیابید.
برای پیدا کردن این مشتری باید balance مشتریهای مختلف را با هم مقایسه کنیم بنابراین ضرب دکارتی به کار می آید و از آنجایی که att نمی تواند در هم ضرب شود پس باید account را در خودش ضرب کنیم و چون نمی توانیم مستقیماً max را بیابیم عبارت کوچکترها به دست آورده و از کل کم کنیم (max هیچ گاه کوچکتر نمی شود)

شماره حساب مشتری با max

$$r_1 = \prod_{account\text{-}\#}(account) - \prod_{account.account\text{-}number}\left(\sigma_{account.balance < x.balance}(account \times \rho_x(account))\right)$$

6 ستون دارد

$$\Rightarrow \text{نام مشتری با max موجودی} = \prod_{customer\text{-}name}(r_1 \bowtie depositor)$$

3. عملگر "Division": فرض کنید می خواهیم نام مشتریهایی که در تمام شعب تهران حساب دارند را پیدا کنیم.
(فرض: تهران سه شعبه دارد)

$$\prod_{customer\text{-}name,\ branch\text{-}name}(account \bowtie depositor) \div \prod_{branch\text{-}name}\left(\sigma_{branch\text{-}city = \text{"Tehran"}}(branch)\right)$$

4 ستون دارد

دو ستونش را جدا می کند

← در تقسیم ستونهای مشترک اعمال می شود.

r_1

customer-name	branch-name
علی	ونک
علی	میرداماد
علی	نازی آباد
رضا	ام
پرهام	ونک

÷

r_2

branch-name
ونک
میرداماد
نازی آباد

⇒

customer-name
علی

Subject: اصول طراحی پایگاه داده‌ها
Year. 86 Month. 8 Date. 19 (13)

روش دوم طرح سوال: یک مرجع مصنوعی ایجاد می‌کنیم یعنی فرض می‌کنیم که تمام مشتریها در تمام شعب تهران حساب دارند و آنهایی که جواب نیستند از این مجموعه مرجع کم می‌کنیم.

r_1 → تمام مشتریها با تمام حسابشان
r_2 → تمام شعب تهران

فرض: تمام افراد در تمام شعب تهران حساب دارند

$$\Pi_{customer\text{-}name}(r_1) - \left(\left(\Pi_{customer\text{-}name}(r_1) \times r_2 \right) - r_1 \right)$$

اسم مشتریهایی که جواب نیستند

Assignment: این عملگر درست باشد" ← "در زبان های برنامه نویسی عمل می‌کند
مثلاً
$temp1 \leftarrow \Pi_{R-S}(r_1)$

جلسه هفتم 86.8.19
مثال (P.57) تمام شعب با مشتریهایی را که در Harrison زندگی می‌کنند و در بانک حساب دارند را بیابید.

از صورت مسئله می‌فهمیم که به branch-name و customer-city نیاز داریم که دو schema Customer و Account را به ما می‌دهد و depositor این دو schema را به هم مربوط می‌سازد پس داریم:

Custome (customer-name, customer-street, customer-city*)
account (account-number, branch-name*, balance)
depositor (customer-name, account-number)

از natural join این سه schema 6 ستون خواهیم داشت که تنها به یک ستون احتیاج داریم. ابتدا به صورت سطری انتخاب می‌کنیم سطرهایی را که customer-city آنها Harrison، و سپس ستون branch-name را چاپ می‌کنیم:

$$\Pi_{branch\text{-}name}\left(\sigma_{customer\text{-}city=\text{"Harrison"}} (customer \bowtie account \bowtie depositor) \right)$$

PAPCO

Subject:
Year. Month. Date. ( )

مثال (P.58) نام مشتریهایی را پیدا کنید که هم وام گرفته‌اند و هم در بانک حساب دارند.

دو schema داریم:
depositor ← مشتریهایی که در بانک حساب دارند.
borrower ← مشتریهایی که از بانک وام گرفته‌اند.
اگر از عملگر ⋈ استفاده نکنیم و ∏ و ∩ می‌توانیم پاسخ صحیح را به دست آوریم اما ⋈ کار را آسان می‌کند

$$\prod_{\text{customer-name}}(\text{depositor}) \cap \prod_{\text{customer-name}}(\text{borrower})$$

$$\rightarrow \prod_{\text{customer-name}}(\text{depositor} \bowtie \text{borrower})$$

نکته: چنانچه دو رابطه r(R) و s(S) هیچ attribute مشترکی نداشته باشند آنگاه:

$R \cap S = \emptyset \Rightarrow r \bowtie s = r \times s$

توجه: عملگر ÷ مناسب Queries است که در آنها "For all" آمده، البته این عبارت همیشه موارد صدق نمی‌کند.

جدول Natural join:

فرض کنید می‌خواهیم برای exp. زیر table رسم کنیم:

$$\prod_{\text{customer-name, branch-name}}(\text{depositor} \bowtie \text{account})$$

$(A_1$, میرباد, 10)		(A_1, آرین)		(A_1, میرباد, 10, آرین)	
(A_2, ونک, 20)	⋈	(A_2, آرین)	=	(A_2, ونک, 20, آرین)	
(A_3, آباران, ∅)		(A_3, رضا)		(A_3, آباران, ∅, رضا)	
مشتری				∏	∏

توجه: δ مقایسه می کند اما از آنجایی که یک عملگر unary است باید دو رابطه در هم ضرب شوند. δ رکورد به رکورد شرط مورد نظر را چک می کند و پس از اتمام رکوردها پاسخ هایی را به ما می دهد.

Extended Relational Model:

عملگرهای جبر رابطه‌ای از چندین روش گسترش می یابند یکی از ساده ترین روش ها استفاده از عملگرهای محاسباتی است. دو روش مهم برای توسعه مدل رابطه ای عبارتند از:

1. aggregate op.
2. outer-join op.

Generalized projection:

$\Pi_{F_1, F_2, \ldots, F_n}(E)$

E: می تواند هر exp. دلخواهی باشد

F_i: هر عبارت جبری می تواند باشد

مثال: می خواهیم ببینیم هر شخص چقدر اعتبار دارد؟ یعنی باید مقدار موجودی شخص را از limit (مقداری که بانک برای هر شخص می پردازد) آن شخص کم کنیم و این کار را با Gen. proj. امکان پذیر می باشد.

$\Pi_{\text{customer-name},\ \text{Limit} - \text{credit-balance}\ \text{as}\ \text{available}}(\text{credit-info})$

«credit-info»

customer-name	Limit	credit-balance
اسماعیل	2000	1500
سنیا	1000	500
علی	1500	800

→

customer-name	available
اسماعیل	500
سنیا	500
علی	700

limit ≥ credit-balance

limit ← کل مقداری که هر شخص می تواند خرج کند

credit_balance ← کل پولی که هر شخص در حساب دارد

Subject:
Year. Month. Date. ( )

همانطور که در مثال می بینیم در Π از عبارت جبری تفاضل استفاده کردیم و حاصل این تفاضل را با کلمه کلیدی as نامگذاری کردیم و پاسخ Π دوستون به نام های customer_name و available دارد.

تفاوت Set و Multiset:
در Set، تکرار معنا ندارد یعنی در مجموعه تکرار مهم نیست اما در Multiset تکرار مهم است:

Set → {1,1,2,2} = {1,2}

Multiset → {1,1,2,2} = {1,1,2,2}

Aggregate Functions

$G_{F(A_1),\dots,F(A_n)}(E)$

مجموعه‌ای از مقادیر را می‌گیرد و یک مقدار را برمی‌گرداند؛ aggregate Fun.

aggregate Fun. در Multiset کار می‌کند یعنی تکرار اهمیت دارد.

- مجموعه {1,1,3,4,4,11} را در نظر بگیرید؛

G_{sum} ← aggregate function sum : مجموع مقادیر را برمی‌گرداند ← 24

G_{avg} ← aggregate function avg : میانگین مقادیر را برمی‌گرداند ← 4

G_{count} ← aggregate function count : تعداد اعضای مجموعه را می‌دهد ← 6

G_{min} ← aggregate function min : کوچکترین عضو مجموعه را می‌دهد ← 1

G_{max} ← aggregate function max : بزرگترین عضو مجموعه را می‌دهد ← 11

$G_{distinct}$ ← aggregate function distinct : از هر تکرار یکی را می‌شمارد (تعداد تمایزها) ← 4

نه همان count است اما در set

مثال: مجموع حقوق که شرکت به کارمندانش می‌دهد

$G_{sum(salary)\ as\ sum\text{-}salary}(employee)$

در جدول employee (schema)

مجموع salary ها را برمی‌گرداند و نام ستون پاسخ را sum_salary می‌گذارد ← پاسخ فقط یک سطر و یک ستون دارد.

P4PCO

Subject: اصول طراحی پایگاه داده‌ها
Year 86 Month 8 Date 19 (15)

« employee »

employee-name	salary	branch-name
یوسف	1,000	میرداماد
طاهره	5,000	ونک
پژمان	12,000	ونک
محمد	6,000	میرداماد

→

sum-salary
24,000

مثال: بیشترین مقدار حقوق را مشخص کنید.

$G_{max(balance),\ account\#}(account)$

مثال (p. 62) مجموع حقوق هر شعبه را مشخص کنید.

$G_{branch\text{-}name}\ sum(salary)\ as\ sum\text{-}salary\ (employee)$

بر حسب نام شعبه گروه بندی می‌کند

employee-name	salary	branch-name
یوسف	1,000	میرداماد
محمد	6000	میرداماد
طاهره	5,000	ونک
پژمان	12,000	ونک

→

branch-name	sum-salary
میرداماد	7,000
ونک	17,000

مثال (p. 64) مجموع حقوق در هر شعبه، max حقوق در هر شعبه؟

$G_{branch\text{-}name}\ sum(salary)\ as\ A,\ max(salary)\ as\ B\ (employee)$

branch-name	A	B
میرداماد	7000	6000
ونک	17000	12000

P4PCO

Subject:
Year. Month. Date. ( )

پس به طور کلی داریم:

$$(G_1, G_2, \ldots, G_n)\ \mathcal{G}^{(E)}_{F_1(A_1), \ldots, F_n(A_n)}$$

G_i ← : att. هایی که در سمت چپ $\mathcal{G}$ می‌آیند باعث گروه بندی می‌شوند یعنی $\mathcal{G}$ ابتدا بر طبق att. G_i ورودی‌ها را گروه بندی کرده و سپس عملیات لازم را انجام می‌دهد

F_i ← : عملیاتی است که باید انجام شود مثل sum و avg و...

A_i ← : نام att. هایی است که عملیات F_i باید روی آن‌ها انجام شود

مثال: مجموع حقوق را بر حسب نام شعبه و نام شهر بدهد یعنی مجموع حقوق هر شعبه، هر شهر را بدهد

employee

city	branch-name	employy-name	salary
تهران	ونک	علی	1000
تهران	ونک	رضا	5,000
اصفهان	بهشتی	زهرا	11,000
اصفهان	بهشتی	یوسف	1,000
تهران	بهشتی	محمد	3,000
تهران	بهشتی	سینا	12,000

باید دو گروه بندی انجام شود یعنی:

G_1 = branch_name

G_2 = city

و دو تا $\mathcal{G}$ داریم:

F_1 = sum

A_1 = salary

$$\text{city, branch-name}\ \mathcal{G}\ \text{sum(salary) as sum-salary}\ (\text{employee})$$

(G_1: city, G_2: branch-name, F_1: sum, A_1: salary)

city	branch-name	sum-salary
تهران	ونک	6,000
تهران	بهشتی	15000
اصفهان	بهشتی	12,000

پاسخ سه ستون دارد که در ستون اول و دوم G_1 و G_2 تکرار می‌دهد و ستون سوم را $F_1(A_1)$.

PAPCO