پایگاه داده ها

طراحی پایگاه داده های رابطه ای

طراحی پایگاه داده‌های رابطه‌ای

- روش‌های مختلف برای طراحی یک پایگاه داده‌های رابطه‌ای
  - استفاده از مدل موجودیت- وابستگی
  - استفاده از فرآیند نرمال‌سازی
- هدف از طراحی یک پایگاه داده‌های رابطه‌ای
  - تولید مجموعه‌ای از شِماهای رابطه‌ای به گونه‌ای که امکان ذخیره‌سازی اطلاعات بدون افزونگیِ غیرضروری فراهم شود
    - طراحی شِماهای رابطه‌ای در یک شکل نرمال مناسب

طراحی پایگاه داده‌های رابطه‌ای

- مثال
  - دو شِمای *loan* و *borrower* را در نظر بگیرید

loan = (*loan_number*, *amount*)

borrower = (*customer_id*, *loan_number*)

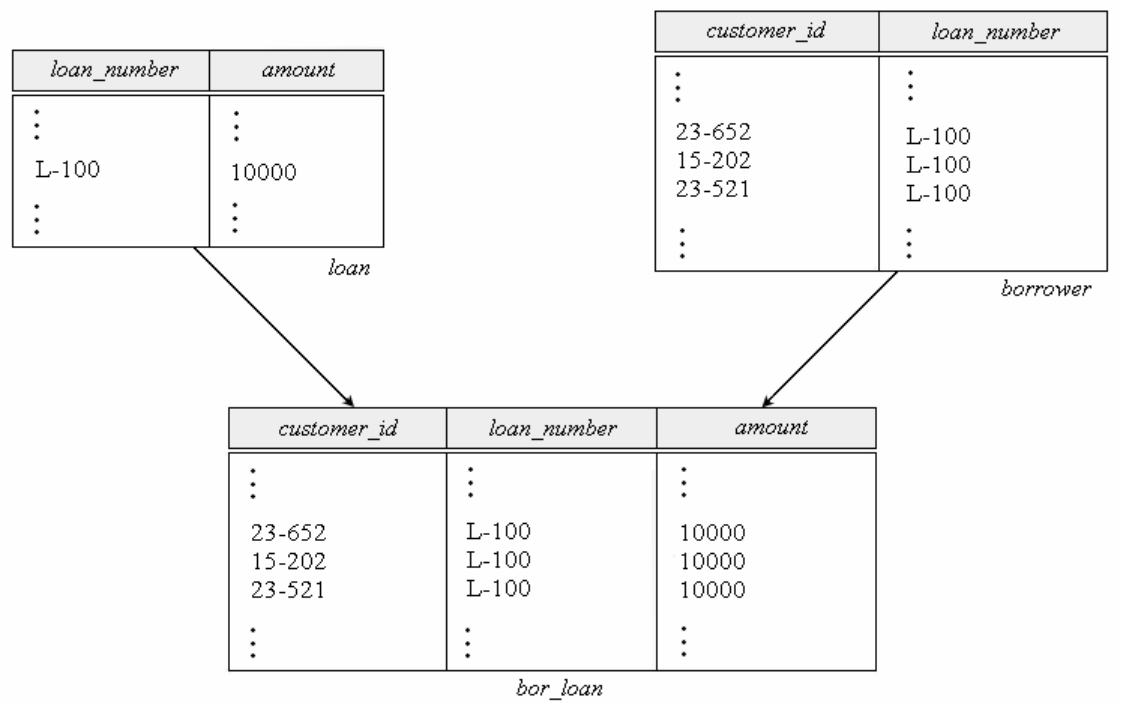
- شِمای *loan* متناظر با مجموعه موجودیت *loan* و شِمای *borrower* متناظر با مجموعه وابستگی *borrower* است
  - مجموعه وابستگی *borrower* چند به چند است
- فرض کنید دو شِمای *loan* و *borrower* با هم ترکیب شوند

bor_loan = (*customer_id*, *loan_number*, *amount*)

- مشکلات
  - افزونگی داده‌ها

طراحی پایگاه داده‌های رابطه‌ای

- رابطه‌های *loan*، *borrower* و *bor_loan*



طراحی پایگاه داده‌های رابطه‌ای

- مثال
  - دو شِمای *loan* و *loan_branch* را در نظر بگیرید

loan = (*loan_number*, *amount*)

loan_branch = (*loan_number*, *branch_name*)

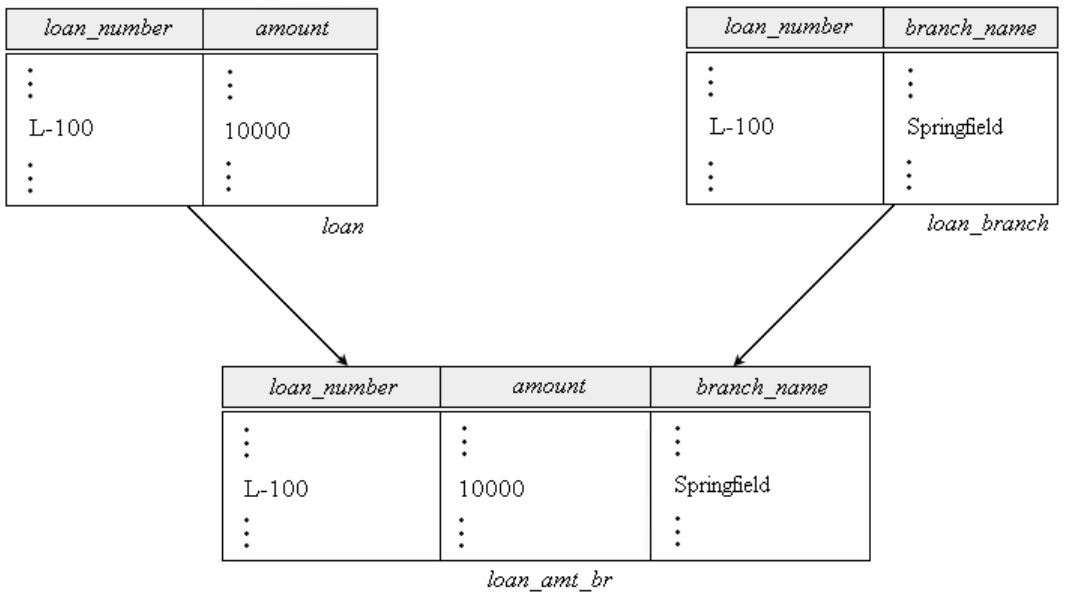
- شِمای *loan* متناظر با مجموعه موجودیت *loan* و شِمای *loan_branch* متناظر با مجموعه وابستگی *loan_branch* است
  - مجموعه وابستگی *loan_branch* چند به یک است

- فرض کنید دو شِمای *loan* و *loan_branch* با هم ترکیب شوند

loan_amt_br = (*loan_number*, *amount*, *branch_name*)

طراحی پایگاه داده‌های رابطه‌ای

- رابطه‌های *loan*، *loan_branch* و *loan_amt_br*
  - مشکل افزونگی داده‌ها وجود ندارد



وابستگی‌های تابعی

- وابستگی تابعی (Functional Dependency)
  - شِمای رابطه‌ای R را در نظر بگیرید
    - فرض کنید $\alpha \subseteq R$ و $\beta \subseteq R$
  - برای هر دو چندگانه t_1 و t_2 در هر رابطه $r(R)$
    - چنانچه مقادیر خصیصه‌های α یکسان باشد، آنگاه مقادیر خصیصه‌های β نیز یکسان باشد
    - در این صورت، وابستگی تابعی $\alpha \rightarrow \beta$ روی شِمای R برقرار است

$$t_1[\alpha] = t_2[\alpha] \Rightarrow t_1[\beta] = t_2[\beta]$$

- وابستگی تابعی $\alpha \rightarrow \beta$ بدیهی نامیده می‌شود اگر $\beta \subseteq \alpha$

وابستگی‌های تابعی

- مثال
  - رابطه r
  - وابستگی‌های تابعی
    - $A \rightarrow C$
    - $A \rightarrow A$
      - وابستگی تابعی بدیهی
    - $AB \rightarrow A$
      - وابستگی تابعی بدیهی

A	B	C	D
a_1	b_1	c_1	d_1
a_1	b_2	c_1	d_2
a_2	b_2	c_2	d_2
a_2	b_3	c_2	d_3
a_3	b_3	c_2	d_4

وابستگی‌های تابعی

- مثال
  - شِمای ***bor_loan***

bor_loan = (<u>***customer_id***, ***loan_number***</u>, ***amount***)

- وابستگی تابعی
  - ***loan_number*** $\rightarrow$ ***amount***
  - برای هر مقدار ***loan_number*** حداکثر یک مقدار ***amount*** وجود دارد

- به دلیل این که ***loan_number*** نمی‌تواند کلید اصلی برای شِمای ***bor_loan*** باشد، میزان یک وام ممکن است چندین بار تکرار شود
  - لازم است ***bor_loan*** به تعدادی شِما تجزیه شود

تجزیه شِماها

- گاهی اوقات ممکن است تجزیه شِماها نامناسب باشد
  - مثال
    - شِمای *employee*

employee = (*employee_id*, *employee_name*, *telephone_number*, *start_date*)

- فرض کنید شِمای *employee* به دو شِمای $employee_1$ و $employee_2$ تجزیه شود

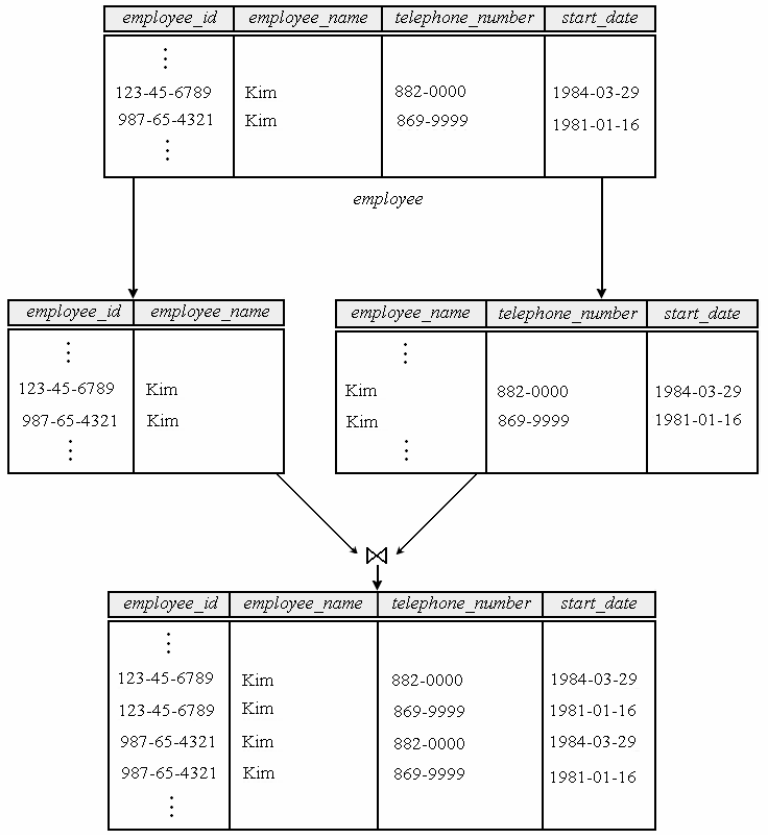
$employee_1$ = (*employee_id*, *employee_name*)

$employee_2$ = (*employee_name*, *telephone_number*, *start_date*)

- امکان از بین رفتن اطلاعات در اثر تجزیه نامناسب فوق وجود دارد

تجزیه شِماها

- تجزیه نامناسب



اولین شکل نرمال (1NF)

- یک دامنه اتمیک است اگر عناصر آن دامنه واحدهای تقسیم‌ناپذیر باشند
- خصیصه‌های مرکب غیراتمیک هستند
  - خصیصه *address*
- شِمای رابطه‌ای R در اولین شکل نرمال (1NF) است اگر دامنه همه خَصیصه‌هایش اتمیک باشند

تجزیه با استفاده از وابستگی‌های تابعی

- در صورتی که یک شِما در شکل نرمال مناسب نباشد، لازم است آن شِما به تعدادی شِمای دیگر تجزیه شود
  - شِماهای حاصل باید در شکل نرمال مناسب باشند
  - در اثر تجزیه نباید اطلاعات از بین بروند
- به منظور تصمیم‌گیری در مورد تجزیه شِماهای رابطه‌ای یک متدولوژی رسمی وجود دارد
  - در این متدولوژی از مفاهیم کلید و وابستگی تابعی استفاده می‌شود

تجزیه با استفاده از وابستگی‌های تابعی

- ابرکلید
  - شِمای رابطه‌ای R را در نظر بگیرید. زیرمجموعه K از R یک ابرکلید برای R نامیده می‌شود اگر $K \rightarrow R$
- مثال
  - شِمای bor_loan

$$bor_loan = (\underline{customer_id, loan_number}, amount)$$

  - زوج خصیصه‌های $(customer_id, loan_number)$ یک ابرکلید برای شِمای bor_loan است

$$customer_id, loan_number \rightarrow customer_id, loan_number, amount$$

یا

$$customer_id, loan_number \rightarrow bor_loan$$

تجزیه با استفاده از وابستگی‌های تابعی

- فرض کنید F مجموعه‌ای از وابستگی‌های تابعی باشد
  - با استفاده از وابستگی‌های تابعی در F ممکن است بتوان وابستگی‌های تابعی دیگری را استنتاج کرد
    - مجموعه این وابستگی‌های تابعی بستار F نامیده می‌شود و با F^+ نمایش داده می‌شود
- مثال
  - شِمای $R = (A, B, C)$ را در نظر بگیرید
  - فرض کنید وابستگی‌های تابعی $A \rightarrow B$ و $B \rightarrow C$ روی شِمای R برقرار باشند
  - می‌توان نتیجه گرفت وابستگی تابعی $A \rightarrow C$ روی شِمای R برقرار است

شکل نرمال BCNF

- فرض کنید مجموعه وابستگی‌های تابعی F روی شِمای رابطه‌ای R برقرار باشند
  - شِمای R در شکل نرمال BCNF است اگر برای همه وابستگی‌های تابعی $\alpha \rightarrow \beta$ در F^+ حداقل یکی از شرایط زیر برقرار باشد
    - $\alpha \rightarrow \beta$ یک وابستگی تابعی بدیهی باشد
      - $\beta \subseteq \alpha$
    - α یک ابرکلید برای شِمای R باشد

- یک طراحی پایگاه داده‌ها در شکل نرمال BCNF است اگر هر کدام از شِماهای رابطه‌ای که آن طراحی را تشکیل می‌دهد در شکل نرمالِ BCNF باشد

شکل نرمال BCNF

- بررسی شِماهای رابطه‌ای برای شکل نرمال BCNF
  - مثال
    - شِمای *bor_loan*

$$bor_loan = (\underline{customer_id, loan_number}, amount)$$

- وابستگی تابعی
  - $loan_number \rightarrow amount$

- *loan_number* یک ابرکلید برای شِمای *bor_loan* نیست
  - شِمای *bor_loan* در شکل نرمال BCNF نیست

- لازم است *bor_loan* به تعدادی شِما تجزیه شود

شکل نرمال BCNF

- فرض کنید شِمای R در شکل نرمال BCNF نباشد
  - حداقل یک وابستگی تابعی غیربدیهی $\alpha \rightarrow \beta$ وجود دارد به گونه‌ای که α یک ابرکلید برای R نیست
  - لازم است شِمای R به دو شِمای دیگر تجزیه شود
    - $(\alpha \cup \beta)$
    - $(R - (\beta - \alpha))$

- مثال
  - تجزیه شِمای *bor_loan*
    - $(\alpha \cup \beta) = (loan_number, amount)$
    - $(R - (\beta - \alpha)) = (customer_id, loan_number)$

شکل نرمال BCNF

- مثال
  - شِمای *lending*

lending = (*branch_name*, *branch_city*, *assets*, *customer_name*, *loan_number*, *amount*)

- مجموعه وابستگی‌های تابعی
  - *branch_name* → *assets branch_city*
  - *loan_number* → *amount branch_name*

- کلید کاندید
  - {*loan_number*, *customer_name*}

- وابستگی تابعی *branch_name* → *assets branch_city* را در نظر بگیرید
  - *branch_name* یک ابرکلید برای شِمای *lending* نیست
  - شِمای *lending* در شکل نرمال BCNF نیست

شکل نرمال BCNF

- لازم است *lending* به تعدادی شِما تجزیه شود

branch = (*branch_name*, *branch_city*, *assets*)

loan_info = (*branch_name*, *customer_name*, *loan_number*, *amount*)

- وابستگی تابعی *branch_name* → *assets branch_city* روی شِمای *branch* برقرار است
  - *branch_name* یک ابرکلید برای شِمای *branch* است
  - شِمای *branch* در شکل نرمال BCNF است
- وابستگی تابعی *loan_number* → *amount branch_name* روی شِمای *loan_info* برقرار است
  - *loan_number* یک ابرکلید برای شِمای *loan_info* نیست
  - شِمای *loan_info* در شکل نرمال BCNF نیست

شکل نرمال BCNF

- لازم است *loan_info* به تعدادی شِما تجزیه شود

loanb = (*loan_number*, *branch_name*, *amount*)

borrower = (*customer_name*, *loan_number*)

- هر دو شِماهای *loanb* و *borrower* در شکل نرمال BCNF هستند

شکل نرمال BCNF و حفظ وابستگی تابعی

- هنگام تجزیه یک شِما با استفاده از الگوریتم تجزیه BCNF امکان از بین رفتن وابستگَی‌های تابعی وجود دارد
  - مثال
    - شِمای *cust_banker_branch*

cust_banker_branch = (*customer_id*, *employee_id*, *branch_name*, *type*)

- فرض کنید هر کارمند تنها در یک شعبه بانک کار کند. همچنین، فرض کنید هر مشتری در هر شعبه بانک حداکثر یک بانکدار شخصی داشته باشد

- مجموعه وابستگی‌های تابعی
  - $employee_id \rightarrow branch_name$
  - $customer_id, branch_name \rightarrow employee_id$

شکل نرمال BCNF و حفظ وابستگی تابعی

- وابستگی تابعی $employee_id \rightarrow branch_name$ را در نظر بگیرید
  - $employee_id$ یک ابرکلید برای شِمای $cust_banker_branch$ نیست
  - شِمای $cust_banker_branch$ در شکل نرمال BCNF نیست
- لازم است $cust_banker_branch$ به تعدادی شِما تجزیه شود

(*employee_id*, *branch_name*)
(*customer_id*, *employee_id*, *type*)

- هر دو شِمای فوق در شکل نرمال BCNF هستند
- با استفاده از طراحی BCNF فوق امکان اعمال قید زیر وجود ندارد
  - هر مشتری در هر شعبه بانک حداکثر یک بانکدار شخصی داشته باشد
  - وابستگی تابعی
    - $customer_id, branch_name \rightarrow employee_id$
- در طراحی BCNF فوق هیچ شِمایی وجود ندارد که شامل همه خصیصه‌های ظاهر شده در وابستگی تابعی فوق باشد

شکل نرمال BCNF و حفظ وابستگی تابعی

- طراحی BCNF فوق حافظ وابستگی‌های تابعی نیست

- همیشه امکان رسیدن به شکل نرمال BCNF و حفظ وابستگی‌های تابعی با هم وجود ندارد
  - لازم است شکل نرمال ضعیف‌تری در نظر گرفته شود
    - سومین شکل نرمال (3NF)

سومین شکل نرمال (3NF)

- فرض کنید مجموعه وابستگی‌های تابعی F روی شِمای رابطه‌ای R برقرار باشند
  - شِمای R در سومین شکل نرمال است اگر برای همه وابستگی‌های تابعی $\alpha \rightarrow \beta$ در F^+ حداقل یکی از شرایط زیر برقرار باشد
    - $\alpha \rightarrow \beta$ یک وابستگی تابعی بدیهی باشد
    - α یک ابرکلید برای شِمای R باشد
    - هر خصیصه A در $\beta - \alpha$ جزیی از یکی از کلیدهای کاندید شِمای R باشد
- اگر یک شِمای رابطه‌ای در شکل نرمال BCNF باشد، آنگاه آن شِمای رابطَه‌ای در سومین شکل نرمال است

سومین شکل نرمال (3NF)

- کلید کاندید
  - شمای رابطه‌ای R را در نظر بگیرید. زیرمجموعه K از R یک کلید کاندید برای R نامیده می‌شود اگر
    - $K \rightarrow R$
    - زیرمجموعه‌ای مانند $\alpha \subset K$ وجود نداشته باشد به گونه‌ای که $\alpha \rightarrow R$

سومین شکل نرمال (3NF)

- بررسی شِماهای رابطه‌ای برای سومین شکل نرمال
  - مثال
    - شِمای *cust_banker_branch*

cust_banker_branch = (*customer_id*, *employee_id*, *branch_name*, *type*)

- مجموعه وابستگی‌های تابعی
  - *employee_id* → *branch_name*
  - *customer_id*, *branch_name* → *employee_id*
- کلیدهای کاندید
  - {*customer_id*, *employee_id*}
  - {*customer_id*, *branch_name*}

سومین شکل نرمال (3NF)

- وابستگی تابعی $employee_id \rightarrow branch_name$ را در نظر بگیرید
  - $\alpha = employee_id$، $\beta = branch_name$ و $\beta - \alpha = branch_name$
  - $branch_name$ جزیی از یکی از کلیدهای کاندید شِمای $cust_banker_branch$ است
- شِمای $cust_banker_branch$ در سومین شکل نرمال است

- **در عمل شِماهای رابطه‌ای باید در شکل نرمال BCNF باشند، مگر این که با اَین کار وابستگی‌های تابعی از بین بروند**
  - در این صورت همان سومین شکل نرمال کفایت می‌کند

بستار یک مجموعه از وابستگی‌های تابعی

- فرض کنید F مجموعه‌ای از وابستگی‌های تابعی باشد
  - بستار F مجموعه همه وابستگی‌های تابعی است که می‌توان آن‌ها را با استفاده از وابستگی‌های تابعی در F استنتاج کرد
    - بستار F با F^+ نمایش داده می‌شود
- برای محاسبه بستار F می‌توان از قوانین آرمسترانگ استفاده کرد
  - قانون انعکاسی
    - اگر α مجموعه‌ای از خصیصه‌ها و $\beta \subseteq \alpha$ باشد، آنگاه $\alpha \rightarrow \beta$ برقرار است
  - قانون افزایشی
    - اگر $\alpha \rightarrow \beta$ برقرار باشد و γ مجموعه‌ای از خصیصه‌ها باشد، آنگاه $\gamma\alpha \rightarrow \gamma\beta$ برقرار است

بستار یک مجموعه از وابستگی‌های تابعی

- قانون تعدی
  - اگر $\alpha \to \beta$ و $\beta \to \gamma$ برقرار باشد، آنگاه $\alpha \to \gamma$ برقرار است

نکته: در قوانین فوق $\alpha \cup \beta$ با $\alpha\beta$ نمایش داده شده است

- با استفاده از قوانین آرمسترانگ می‌توان قوانین زیر را نتیجه گرفت
  - قانون اجتماع
    - اگر $\alpha \to \beta$ و $\alpha \to \gamma$ برقرار باشد، آنگاه $\alpha \to \beta\gamma$ برقرار است
  - قانون تجزیه
    - اگر $\alpha \to \beta\gamma$ برقرار باشد، آنگاه $\alpha \to \beta$ و $\alpha \to \gamma$ برقرار است
  - قانون شبه تعدی
    - اگر $\alpha \to \beta$ و $\beta\gamma \to \delta$ برقرار باشد، آنگاه $\alpha\gamma \to \delta$ برقرار است

بستار یک مجموعه از وابستگی‌های تابعی

- مثال
  - شِمای $R = (A, B, C, G, H, I)$
  - مجموعه وابستگی‌های تابعی
    - $F = \{A \rightarrow B, A \rightarrow C, CG \rightarrow H, CG \rightarrow I, B \rightarrow H\}$

- تعدادی از اعضای F^+
  - $A \rightarrow H$
  - $CG \rightarrow HI$
  - $AG \rightarrow I$

بستار یک مجموعه از خصیصه‌ها

- فرض کنید α مجموعه‌ای از خصیصه‌ها و F مجموعه‌ای از وابستگی‌های تابعی باشد
  - خصیصه B به صورت تابعی توسط α تعیین می‌شود اگر $\alpha \rightarrow B$
  - مجموعه همه خصیصه‌هایی که به صورت تابعی توسط α تعیین می‌شوند، بستار α نامیده می‌شود
    - بستار α با α^+ نمایش داده می‌شود
- از بستار مجموعه خصیصه‌های α می‌توان برای بررسی ابرکلید بودن α استفاده کرد

بستار یک مجموعه از خصیصه‌ها

- الگوریتم محاسبه α^+

result := α;
while (changes to *result*) **do**
    for each functional dependency $\beta \rightarrow \gamma$ **in** F **do**
    begin
        if $\beta \subseteq$ *result* **then** *result* := *result* $\cup$ γ;
    end

بستار یک مجموعه از خصیصه‌ها

- مثال
  - شِمای $R = (A, B, C, G, H, I)$
  - مجموعه وابستگی‌های تابعی
    - $F = \{A \rightarrow B, A \rightarrow C, CG \rightarrow H, CG \rightarrow I, B \rightarrow H\}$
  - محاسبه $(AG)^+$
    - $result = AG$
    - $result = ABG$ ($A \rightarrow B$ and $A \subseteq AG$)
    - $result = ABCG$ ($A \rightarrow C$ and $A \subseteq ABG$)
    - $result = ABCGH$ ($CG \rightarrow H$ and $CG \subseteq ABCG$)
    - $result = ABCGHI$ ($CG \rightarrow I$ and $CG \subseteq ABCGH$)
  - بررسی ابرکلید بودن AG
    - $R \subseteq (AG)^+$
    - بنابراین $AG \rightarrow R$

بستار یک مجموعه از خصیصه‌ها

- بررسی ابرکلید بودن مجموعه خصیصه‌های α
  - ابتدا لازم است α^+ محاسبه شود، سپس بررسی شود که $R \subseteq \alpha^+$
- بررسی برقرار بودن وابستگی تابعی $\alpha \rightarrow \beta$
  - ابتدا لازم است α^+ محاسبه شود، سپس بررسی شود که $\beta \subseteq \alpha^+$