

# طراحی بانک های اطلاعات توزیع شده:

طراحی یک سیستم بانک اطلاعات نامتمرکز شامل:

- (1) طراحی شبکه
- (2) توزیع داده ها و برنامه ها در سایتهای این شبکه

دو استراتژی کلی در طراحی:

- (1) طراحی بالا به پایین: ابتدا شمای سراسری بعد شمای محلی
- (2) طراحی پایین به بالا: ابتدا شمای محلی بعد شمای سراسری

# توزیع (Distribution):

توزیع در بانک اطلاعات نامتمرکز رابطه ای شامل دو بخش است:

(1) تقسیم : تقسیم رابطه ها به بخش های کوچکتر

(2) تخصیص : تخصیص این بخش ها به سایت ها

رابطه از لحاظ توزیع داده ها واحد مناسبی نمی باشد، بنابراین باید رابطه ها را به بخش های کوچکتری تقسیم کرد و سپس آن را در سایتهای مختلف توزیع کرد.

# اطلاعات مورد نیاز در توزیع:

- (1) اطلاعات بانک اطلاعات
- (2) اطلاعات برنامه کاربردی
- (3) اطلاعات شبکه ارتباطی
- (4) اطلاعات سیستم کامپیوتری

## تقسیم:

مزیت تقسیم رابطه ها به بخش های کوچکتر:  
افزایش درجه همروندی و گذردهی سیستم

زیرا:

- (1) امکان اجرای همروند چندین تراکنش
- (2) تجزیه احتمالی یک پرس وجوبه مجموعه ای از زیرپرس وجوها  
و اجرای موازی آنها روی تکه های مختلف (همروندی درون پرس و  
جو)

# معایب تقسیم رابطه ها:

- (1) اگر در برنامه کاربردی نیازمندیهایی داشته باشیم که با تقسیم رابطه ها در تضاد باشد، کارائی پایین می آید.
- (2) کنترل معنایی داده است که با تقسیم کردن، صفاتی که در یک وابستگی قرار می گیرند ممکن است به چند تکه تقسیم شوند و کنترل جامعیت آنها پیچیده تر است.

# انواع تقسیم:

- (1) تقسیم سطری (horizontal)
- (2) تقسیم ستونی (vertical)
- (3) تقسیم ترکیبی (hybrid)

# تقسیم سطری:

هر قسمت، زیر مجموعه ای از تاپل های رابطه است. این نوع تقسیم را به دو شکل می توان انجام داد:

**(1) تقسیم سطری اولیه (primary):** انتخاب سطرها بر اساس شرط هایی که روی همان رابطه تعریف می کنیم، انجام می شود.

**(2) تقسیم سطری مشتق (derived):** رابطه را بر اساس شرط های تعریف شده روی یک رابطه دیگر تقسیم می نمائیم.

## شرط ساده و شرط کمینه:

شرط ساده: اگر  $R(A_1, A_2, \dots, A_n)$  رابطه مورد نظر و  $A_i$  صفتی از دامنه  $D_i$  باشد یک شرط ساده به شکل  $\theta$  value  $P_j:A_i$  است که  $\theta$  عملگر مقایسه ای و  $\theta$  مقدار از دامنه  $D_i$  است.

شرط کمینه: اگر  $P_r = \{P_1, P_2, \dots, P_n\}$  مجموعه شرط های ساده روی رابطه  $R$  باشد، مجموعه شرط های کمینه آن برابر است با:

$$M = \{m_i \mid m_i = P_i \wedge P_j \in P_r, P_j^*\}, 1 \leq j \leq m, 1 \leq i \leq z$$



# تقسیم سطری اولیه، تقسیم ستونی، تقسیم ترکیبی:

تقسیم سطری اولیه: روی رابطه ها فقط عمل گزینش انجام می شود.  
هرتکه  $R_i$  اینگونه تعریف می شود:

$$R_i = \delta F_i(R)$$

$F_i$  می تواند یک شرط ساده یا یک شرط کمینه باشد.

تقسیم ستونی: با تقسیم ستونی رابطه  $R$  به بخش های  $R_i$ ، هر یک از  $R_i$  ها شامل زیر مجموعه ای از صفات رابطه  $R$  به علاوه کلید اصلی  $R$  خواهد بود.

تقسیم ترکیبی: گاهی تقسیم سطری یا تقسیم ستونی به تنهایی برای برنامه های کاربردی کافی نیست و نیاز است پس از تقسیم سطری از تقسیم ستونی استفاده شود.

# شرایط صحت تقسیم:

- (1) **کامل بودن (completeness):** اگر داده های مربوط به رابطه  $R$  به رابطه های  $R_1, \dots, R_n$  تقسیم شوند هر قلم داده ای که در  $R$  وجود دارد، باید در یک یا بیشتر از  $R_i$  ها وجود داشته باشد.
- (2) **قابل بازسازی مجدد (reconstruction):** اگر رابطه  $R$  به رابطه های  $R_1, \dots, R_n$  تقسیم شود. باید بتوان از رابطه های حاصله رابطه اولیه را به دست آورد.
- (3) **عدم افزونگی (disjontitnes):** اگر رابطه  $R$  به رابطه های  $R_1, \dots, R_n$  تقسیم شود، هیچ قلم داده ای نباید در بیش از یک  $R_i$  تکرار گردد.

# تخصیص (Allocation):

نحوه تخصیص بخشهای رابطه تقسیم شده به سایتهای مختلف روی شبکه.

انواع بانک اطلاعات از نظر تکرار:

(1) فاقد تکرار: از هر تکه فقط یک کپی در شبکه است.

(2) دارای تکرار

✓ تکرار کامل: هر تکه در همه سایتهای شبکه وجود دارد.

✓ تکرار جزئی: هر تکه ممکن است در چند سایت تکرار شود.

# اطلاعات مورد نیاز برای تخصیص:

- ✓ اندازه هر تکه
- ✓ تعداد تاپلهای آن تکه
- ✓ فرکانس دسترسی
- ✓ مکان دسترسی
- ✓ نوع دسترسی
- ✓ هزینه ذخیره سازی
- ✓ هزینه پردازش پرس و جو
- ✓ پهنای باند
- ✓ تاخیر و سر بار انجام ارتباطات