## 3. Duomenų bazių projektavimas

- DB projektavimo rezultatas - loginė DB struktūra.
- Ieškoma DB struktūros su geromis savybėmis.
- DB kokybei įvertinti naudojami formalūs metodai.



### *3.1. Pagrindinės reliacinio modelio sąvokos*

**Reliaciniame modelyje** duomenys vaizduojami lentelėmis, kitaip dar vadinamomis **santykiais**.

Santykis (angl. ***relation***) – matematikos terminas – 2D lentelė.

Lentelės stulpelis dar vadinamas **atributu.**

Santykio eilutės – **kortežai**.

Eilučių tvarka santykyje yra neapibrėžta.

Visų atributo reikšmių aibė – **atributo sritimi (domenu)**. Tuščia reikšmė – `NULL`.



Atributų aibė, vienareikšmiškai apibrėžianti kiekvieną lentelės kortežą (eilutę), vadinama lentelės **viršrakčiu** (**superraktu**).

Lentelės **raktas** – viršraktis, iš kurio pašalinus bet kurį atributą, jis nustoja būti viršrakčiu.

**Lentelės *L* raktas** – atributų aibės poaibis *K*, toks kad:
1) (**vienareikšmiška identifikacija**) rakto atributų reikšmės vienareikšmiai apibrėžia visų lentelės *L* atributų reikšmes;
2) (**pertekliaus nebuvimas**) joks aibės *K* poaibis neturi vienareikšmės identifikacijos savybės.



*Vykdytojai*:

| | |
|---|---|
| {*Nr*, *Pavardė*} | – **viršraktis**, |
| {*Nr*, *Pavardė*, *Kategorija*} | – **viršraktis**, |
| {*Nr*, *Kategorija*} | – **viršraktis**, |
| {*Nr*} | – **raktas**. |

Lentelė gali turėti kelis raktus.
Visi raktai vadinami (**potencialiais**, **galimais**) **raktais**.
Vienas raktas paskelbiamas **pirminiu raktu**.



**Lentelės reliacinė schema** (struktūra, aprašas) - lentelės ir visų jos stulpelių pavadinimai, su pažymėtu pirminiu raktu:

*Vykdytojai* (*Nr*, *Pavardė, Kvalifikacija, Kategorija, Išsilavinimas*)

*Projektai*(*Nr*, *Pavadinimas, Svarba, Pradžia, Trukmė*)

*Vykdymas* (*Projektas*, *Vykdytojas*, *Statusas, Valandos*)



**Išoriniu (svetimuoju) raktu** vadinamas lentelės atributų rinkinys, kuris kitoje ar net toje pačioje lentelėje yra pirminis raktas.

*Vykdymas.Vykdytojas* – išorinis raktas, kurį atitinkantis stulpelis *Nr* lentelėje *Vykdytojas* yra pirminis raktas.

Išorinis raktas apibrėžiamas atributais ir kitos lentelės pavadinimu.

*Vykdymas.Projektas* – išorinis raktas, nukreipiantis į *Projektas*.



**DB bazės reliacinė schema** - visų jos lentelių reliacinių schemų rinkinys kartu su lentelių išoriniais raktais.

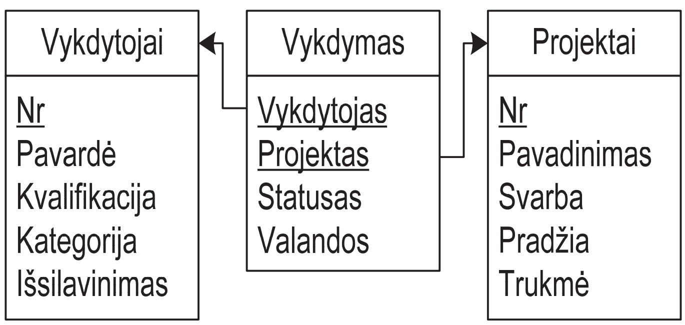
**DB** *Darbai* **schema**:

*Vykdytojai* (*Nr*, *Pavardė, Kvalifikacija, Kategorija, Išsilavinimas*);
*Projektai*(*Nr*, *Pavadinimas, Svarba, Pradžia,Trukmė*);
*Vykdymas* (*Projektas*, *Vykdytojas*, *Statusas,Valandos*);

**Išoriniai raktai**: *Projektas* nukreipia į *Projektai*,
*Vykdytojas* nukreipia į *Vykdytojai*.



### *Grafinis DB schemos vaizdavimas*



### *3.2. Reliacinė algebra*

Tarkime, $A_1, A_2,..., A_n$ yra lentelės $L$ atributų (stulpelių) aibė.

$L(A_1, A_2,..., A_n)$ – **lentelės schema**

Jei $R = \{A_1, A_2,..., A_n\}$, tai $L(R)$ – schema.

**Lentelė** $l$ su schema $L(A_1, A_2,..., A_n)$ - sutvarkytų reikšmių rinkinių aibė:

$l = \{e_1, e_2,..., e_m\}$

**Lentelė *l* su schema** $L(A_1, A_2,..., A_n)$ žymima

$l(A_1, A_2,..., A_n)$



Kiekviena **eilutė** $e_j$ yra sutvarkytas reikšmių rinkinys

$e_j = <a_1, a_2,..., a_n> \in l,\ a_i \in dom(A_i)$

$e(A_i)$ - atributo $A_i$ **reikšmė** eilutėje $e$

$e(A)$, atributų $A \subseteq R$ reikšmių rinkinys eilutėje $e$

$V \subseteq R$ yra $l(R) = \{e_1, e_2,..., e_m\}$ **viršraktis**, jei

$\forall\ i, j = 1,...,m,\ i \neq j: e_i(V) \neq e_j(V)$

Lentelę galima apibrėžti **Dekarto sandaugos poaibiu**

$l(A_1, A_2,..., A_n) \subseteq (dom(A_1) \times dom(A_2) \times ... \times dom(A_n))$

**Schema** $L(R)$ ir konkretus jos **turinys** $l(R)$ **ne tas pat**.

$l(R)$ – lentelės su schema $L(R)$ konkretus turinys.



**Pavyzdys**:

Tarkime: $dom(A) = \{a_1, a_2, a_3\}$ ir $dom(B) = \{b_1, b_2\}$.

Tuomet: $dom(A) \times dom(B) = \{<a_1, b_1>, <a_1, b_2>, <a_2, b_1>, <a_2, b_2>, <a_3, b_1>, <a_3, b_2>\}$

Lentelės su schema $T(A, B)$ turinių, pavyzdžiai:

| $A$ | $B$ |
|---|---|
| $a_1$ | $b_1$ |
| $a_1$ | $b_2$ |
| $a_2$ | $b_1$ |

| $A$ | $B$ |
|---|---|
| $a_1$ | $b_2$ |
| $a_2$ | $b_1$ |

| $A$ | $B$ |
|---|---|
| $a_1$ | $b_2$ |



Operacijų su lentelėmis sistema - **reliacinė algebra**.

**Reliacine algebra** - formali operacijų kalba (sistema), kuria iš lentelių, nekeičiant jų turinio, galima gauti kitas lenteles.

Operacijos **rezultatas** - **lentelė**.

Galima konstruoti **reliacinės algebros reiškinius**.

Reliacinės algebros **operacijų grupės**:

1) matematinės **aibių operacijos**: $\cup, \cap, -, \times$.
2) specifinės DB operacijos:
   - **išrinkimas** (angl. *selection*) ($\sigma$)
   - **projekcija** (angl. *projection*) ($\pi$)
   - **jungimas** (angl. *join*) ($\bowtie$) ir kitos.



### *3.2.1. Pagrindinės operacijos su lentelėmis*

**Išrinkimas** – unarinė operacija**.**

Tarkime, $l = l(R)$ - lentelė, $A \in R,\ a \in dom(A)$

„iš $l$ išrinkti visas eilutes, kuriose atributo $A$ reikšmė yra lygi $a$“:

$\sigma_{A=a}(l) = l'(R) = \{e \in l : e(A) = a\}$

$\sigma_{A=a}(\sigma_{B=b}(l)) = \sigma_{B=b}(\sigma_{A=a}(l))$ - **komutatyvi**

$\sigma_{A=a}(\sigma_{B=b}(l)) = \sigma_{A=a}(\{e \in l : e(B) = b\}) =$

$= \{e' \in \{e \in l : e(B) = b\} : e'(A) = a\} =$

$= \{e \in l : e(A) = a,\ e(B) = b\} =$

$= \{e' \in \{e \in l : e(A) = a\} : e'(B) = b\} = \sigma_{B=b}(\sigma_{A=a}(l))$



Dėl komutyvumo, išrinkimo tvarka yra nesvarbi, todėl

$\sigma_{A=a} \circ \sigma_{B=b} \equiv \sigma_{A=a,B=b}$

$\sigma_{A=a}$ – distributyvi aibių operacijų ($\cup, \cap, -$) atžvilgiu:

$\sigma_{A=a}(l_1 \cup l_2) = \sigma_{A=a}(l_1) \cup \sigma_{A=a}(l_2)$

$\sigma_{A=a}(l_1 \cap l_2) = \sigma_{A=a}(l_1) \cap \sigma_{A=a}(l_2)$

$\sigma_{A=a}(l_1 - l_2) = \sigma_{A=a}(l_1) - \sigma_{A=a}(l_2)$



**Projekcija** – unarinė operacija.

$l = l(R)$ projekcija aibėje $A$, $\pi_A(l)$,

$A \subseteq R$, yra lentelė $l'(A)$, kuri gaunama iš $l(R)$ išbraukiant stulpelius $R–A$,

$\pi_A(l) = l'(A) = \{e(A) : e \in l\}$

Jei $A \subseteq B \subseteq R$, tai lentelei $l(R)$ galioja:

$\pi_A(\pi_B(l)) = \pi_A(l)$



**Projekcija** yra **komutatyvi išrinkimo** atžvilgiu.

Jei $A \in R,\ a \in dom(A),\ B \subseteq R$ ir $l(R)$ - lentelė, tai

$\pi_B(\sigma_{A=a}(l)) = \pi_B(\{e \in l : e(A) = a\}) =$

$= \{e'(B) : e' \in \{e \in l : e(A) = a\}\} =$

$= \{e(B) : e \in l \text{ ir } e(A) = a\} =$

$= \sigma_{A=a}(\{e(B) : e \in l\})$

$= \sigma_{A=a}(\pi_B(l))$

**Pavyzdys**:

$l(A, B)$

| $A$ | $B$ |
|---|---|
| $a_1$ | $b_1$ |
| $a_1$ | $b_2$ |
| $a_2$ | $b_1$ |

$\sigma_{A=a1}(l)$

| $A$ | $B$ |
|---|---|
| $a_1$ | $b_1$ |
| $a_1$ | $b_2$ |

$\pi_B(l)$

| $B$ |
|---|
| $b_1$ |
| $b_2$ |

$\sigma_{B=b1}(l)$

| $A$ | $B$ |
|---|---|
| $a_1$ | $b_1$ |
| $a_2$ | $b_1$ |

$\pi_A(l)$

| $A$ |
|---|
| $a_1$ |
| $a_2$ |

$\pi_B(\sigma_{A=a1}(l))$

| $B$ |
|---|
| $b_1$ |
| $b_2$ |



**Jungimas** – binarinė operacija.

Kombinuojamos dvi lentelės.

$l_1(R_1)$ ir $l_2(R_2)$ **junginys** $l_1 \bowtie l_2$ yra lentelė $l_3(R_3)$:

$R_3 = R_1 \cup R_2$

$l_1 \bowtie l_2 = l_3(R_3) =$

$\{e : \exists\, e_1, e_2\ (e_1 \in l_1,\ e_2 \in l_2,\ e(R_1) = e_1,\ e(R_2) = e_2,$
$e_1(R_1 \cap R_2) = e_2(R_1 \cap R_2))\ \}$

Jei $R_1 \cap R_2 = \varnothing$, tai jungimas yra ekvivalentiškas Dekarto sandaugai.



### *3.2.2. Jungimo savybės*

**Jungimu** galima išreikšti išrinkimo operaciją $\sigma_{A=a}(l)$.

Tarkime, $l'(A)$ yra 1 eilutės lentelė, $e(A) = a$.

$l \bowtie l' = \{e : \exists\, e_1, e_2\ (e_1 \in l,\ e_2 \in l',\ e(R) = e_1,\ e(A) = e_2)\}$

$= \{e : \exists\, e_1\ (e_1 \in l,\ e(R) = e_1,\ e(A) = a)\}$

$= \sigma_{A=a}(l)$

Pvz.:

$l(A, B)$

| $A$ | $B$ |
|---|---|
| $a_1$ | $b_1$ |
| $a_1$ | $b_2$ |
| $a_2$ | $b_1$ |

$\sigma_{A=a1}(l)$

| $A$ | $B$ |
|---|---|
| $a_1$ | $b_1$ |
| $a_1$ | $b_2$ |

$l'(A)$

| $A$ |
|---|
| $a_1$ |

$l \bowtie l'$

| $A$ | $B$ |
|---|---|
| $a_1$ | $b_1$ |
| $a_1$ | $b_2$ |

Tapačios



$(l_1 \bowtie l_2) \bowtie l_3 = l_1 \bowtie (l_2 \bowtie l_3)$

– dėl simetrijos apibrėžime

$l_1$

| $A$ | $B$ |
|---|---|
| $a_1$ | $b_1$ |
| $a_1$ | $b_2$ |
| $a_2$ | $b_1$ |

$l_2$

| $B$ | $C$ |
|---|---|
| $b_1$ | $c_2$ |
| $b_2$ | $c_1$ |

$l_3$

| $A$ | $C$ |
|---|---|
| $a_1$ | $c_2$ |
| $a_2$ | $c_2$ |

$l_1 \bowtie l_2 \bowtie l_3$

| $A$ | $B$ | $C$ |
|---|---|---|
| $a_1$ | $b_1$ | $c_2$ |
| $a_2$ | $b_1$ | $c_2$ |



$l_1$ eilutė $<a_1, b_2>$ ir $l_2$ eilutė $<b_2, c_1>$ liko nesujungtos.

$l_1, l_2, \ldots, l_m$ – **visiškai sujungiamos**, jei kiekviena kiekvienos lentelės eilutė yra lentelių junginio konkrečios eilutės dalis.

Pvz., jei $l_3$ papildytume eilute $<a_1, c_1>$, tai $l_1, l_2, l_3$ taptų visiškai sujungiamomis.

$l_1 \bowtie l_2 \bowtie l_3$

| $A$ | $B$ | $C$ |
|---|---|---|
| $a_1$ | $b_1$ | $c_2$ |
| $a_1$ | $b_2$ | $c_1$ |
| $a_2$ | $b_1$ | $c_2$ |



$l_1$

| $A$ | $B$ |
|---|---|
| $a_1$ | $b_1$ |
| $a_1$ | $b_2$ |
| $a_2$ | $b_1$ |

$l_2$

| $B$ | $C$ |
|---|---|
| $b_1$ | $c_2$ |
| $b_2$ | $c_1$ |

$l_3$

| $A$ | $C$ |
|---|---|
| $a_1$ | $c_2$ |
| $a_2$ | $c_2$ |
| $a_1$ | $c_1$ |

$l_1 \bowtie l_2 \bowtie l_3$

| $A$ | $B$ | $C$ |
|---|---|---|
| $a_1$ | $b_1$ | $c_2$ |
| $a_1$ | $b_2$ | $c_1$ |
| $a_2$ | $b_1$ | $c_2$ |



Bet kurioms lentelėms $l_1(R_1)$ ir $l_2(R_2)$ galioja

$$\pi_{R_1}(l_1 \bowtie l_2) \subseteq l_1$$

ir

$$\pi_{R_2}(l_1 \bowtie l_2) \subseteq l_2$$

Čia $\subseteq$ tampa $=$, kai

$\forall e_1 \in l_1, \exists e_2 \in l_2 : e_1(R_1 \cap R_2) = e_2(R_1 \cap R_2)$

Lygybei nėra būtina, kad $l_1$ ir $l_2$ būtų visiškai sujungiamos.



Čia $l_2$ ir $l_3$ nėra visiškai sujungiamos, nes negalima sujungti $<b_2, c_1> \in l_2$

$l_2$

| $B$ | $C$ |
|---|---|
| $b_1$ | $c_2$ |
| $b_2$ | $c_1$ |

$l_3$

| $A$ | $C$ |
|---|---|
| $a_1$ | $c_2$ |
| $a_2$ | $c_2$ |

$l_2 \bowtie l_3$

| $A$ | $B$ | $C$ |
|---|---|---|
| $a_1$ | $b_1$ | $c_2$ |
| $a_2$ | $b_1$ | $c_2$ |

$l_3 = \pi_{AC}(l_2 \bowtie l_3)$, $\pi_{BC}(l_2 \bowtie l_3) \subset l_2$

$\pi_{AC}(l_2 \bowtie l_3)$

| $A$ | $C$ |
|---|---|
| $a_1$ | $c_2$ |
| $a_2$ | $c_2$ |

$\pi_{BC}(l_2 \bowtie l_3)$

| $B$ | $C$ |
|---|---|
| $b_1$ | $c_2$ |

Tarkime $l_3(R_1 \cup R_2)$ yra lentelė, o $l_1(R_1)$ ir $l_2(R_2)$ – jos projekcijos,

$$l_1 = \pi_{R_1}(l_3), \;\; l_2 = \pi_{R_2}(l_3)$$

Jei $e \in l_3$, tai $e(R_1) \in l_1$, $e(R_2) \in l_2 \;\Rightarrow\; l_3 \subseteq l_1 \bowtie l_2$

$l_3$

| A | B | C |
|---|---|---|
| $a_1$ | $b_1$ | $c_1$ |
| $a_2$ | $b_1$ | $c_1$ |
| $a_2$ | $b_1$ | $c_2$ |

$l_1$

| A | B |
|---|---|
| $a_1$ | $b_1$ |
| $a_2$ | $b_1$ |

$l_2$

| B | C |
|---|---|
| $b_1$ | $c_1$ |
| $b_1$ | $c_2$ |

$l_1 \bowtie l_2$

| A | B | C |
|---|---|---|
| $a_1$ | $b_1$ | $c_1$ |
| $a_1$ | $b_1$ | $c_2$ |
| $a_2$ | $b_1$ | $c_1$ |
| $a_2$ | $b_1$ | $c_2$ |



**Jei** $l_3 = l_1 \bowtie l_2$, **tai** $l_3$ skaidymas į $l_1$ ir $l_2$ - **dekompozicija be praradimo.**

Taip skaidant **neprarandami jokie duomenys**.

$l_3$

| A | B | C |
|---|---|---|
| $a_1$ | $b_1$ | $c_1$ |
| $a_2$ | $b_1$ | $c_1$ |
| $a_2$ | $b_1$ | $c_2$ |

$l_1$

| A | C |
|---|---|
| $a_1$ | $c_1$ |
| $a_2$ | $c_1$ |
| $a_2$ | $c_2$ |

$l_2$

| B | C |
|---|---|
| $b_1$ | $c_1$ |
| $b_1$ | $c_2$ |

$l_1 \bowtie l_2$

| A | B | C |
|---|---|---|
| $a_1$ | $b_1$ | $c_1$ |
| $a_2$ | $b_1$ | $c_1$ |
| $a_2$ | $b_1$ | $c_2$ |



Operacijų aibė $\{\sigma, \pi, \cup, -, \times\}$ yra **pilnoji**.

**Pvz.**, $\cap$ galima išreikšti **per** $\cup$ **ir** $-$

$$l_1 \cap l_2 = (l_1 \cup l_2) - ((l_1 - l_2) \cup (l_2 - l_1))$$

$l_1(R_1)$ ir $l_2(R_2)$ junginį $l_1 \bowtie l_2$ galima išreikšti per $\sigma$, $\pi$ ir $\times$

$$l_1 \bowtie l_2 = \pi_{R_1 \cup R_2}(\sigma_{l_1.A_1 = l_2.A_1, \ldots, l_1.A_N = l_2.A_N}(l_1 \times l_2)),$$

$$\{A_1, A_2, \ldots, A_N\} = R_1 \cap R_2$$



### *3.2.3. Palyginimo apibendrinimas operacijose*

Domeno reikšmėms palyginti naudojome '=' operaciją.

Reliacinėje teorijoje bet kurias domeno reikšmes galima įvertinti, ar jos tarpusavyje lygios (=), ar nelygios ($\neq$)

Dažnai domeno 2 reikšmes galima įvertinti,

**kuri iš jų yra didesnė.**

Daugelyje domenų naudojama: **=, ≠, <, ≤, ≥, >**

Šią operacijų **aibę** žymėsime **Θ**.

Bendruoju atveju gali būti lyginamos skirtingų domenų reikšmės.



Jei $\theta \in \Theta$, o $A$ ir $B$ – atributai,

tai $A$ ir $B$ – vadinsime $\theta$–palyginamais,

jei tik $\theta$ yra apibrėžta aibėje $dom(A) \times dom(B)$.

Išrinkimą ir jungimą galima išplėsti kitoms palyginimo operacijoms.

Jei $l = l(R)$ - lentelė, $A, B \in R$, $a \in dom(B)$, $\theta \in \Theta$, $A$ ir $B$ yra $\theta$–palyginami,

tai $\sigma_{A\theta a}(l)$ žymi $\theta$ **– išrinkimo operaciją**

$$\sigma_{A\theta a}(l) = \{e \in l : e(A)\ \theta\ a\}$$

Vietoje atributo lyginimo su reikšme, galima lyginti du atributus: $\sigma_{A\theta B}(l) = \{e \in l : e(A)\ \theta\ e(B)\}$



Išrinkimą galima dar **apibendrinti**, leidžiant sąlygoje naudoti logines operacijas bei skliaustelius, pvz.

$$\sigma_{A \le B \wedge (A=a \vee B \neq b)}(l)$$

Jungiant lenteles galima naudoti bet kurią palyginimo operaciją $\theta \in \Theta$.

Dviejų lentelių $l_1(R_1)$ ir $l_2(R_2)$ **$\theta$ – junginiu** $l_1 \bowtie_\theta l_2$ vadinama lentelė, sudaryta iš stulpelių $R_3 = R_1 \cup R_2$, ir kurios eilutės apibrėžiamos taip

$$l_1 \bowtie_\theta l_2 = \{e : \exists\, e_1 \in l_1, \exists\, e_2 \in l_2, e(R_1) = e_1, e(R_2) = e_2,$$
$$e_1(R_1 \cap R_2)\ \theta\ e_2(R_1 \cap R_2)\ \}$$



Tiek $\theta$ – išrinkimas, tiek ir apibendrintasis išrinkimas yra plačiai naudojami, $\theta$ – jungimas - rečiau.

Visos anksčiau apibrėžtos išrinkimo ir jungimo operacijų savybės būdingos ir jų pateiktiesiems apibendrinimams.



### *Reliacinės operacijos SQL kalboje*

Visas pagrindines reliacines operacijas realizuoja SQL sakinys **SELECT**.

Lentelės $l(R)$ **išrinkimo** operacija $\sigma_{A=a}(l)$:

```
SELECT * FROM l WHERE A  = a
```

$\sigma_{A \le B \wedge (A=a \vee B \neq b)}(l)$ galima užrašyti taip

```
SELECT * FROM l
WHERE A <= B AND (A = a OR B <> b)
```

Lentelės $l(R)$ **projekciją** $\pi_A(l)$ galima išreikšti sakiniu

**SELECT DISTINCT** *A* **FROM** *L*

Kadangi reliacinėje teorijoje lentelėje vienodų eilučių nėra, tai projekcijai išreikšti prireikė raktažodžio **DISTINCT**.

$l_1(R_1)$ ir $l_2(R_2)$ **junginys** $l_1 \bowtie l_2$

**SELECT** $R_1 \cup R_2$ **FROM** $l_1$, $l_2$
**WHERE** $l_1.A_1 = l_2.A_1$ **AND** . . . **AND** $l_1.A_N = l_2.A_N$
čia $(A_1, A_2, \ldots, A_N) = R_1 \cap R_2$



**θ–junginys** $l_1 \bowtie_\theta l_2$ išreiškiamas panašiai.

**Aibių operacijų**: $\cup$, $\cap$, $-$ atitikmenys:

**UNION**, **INTERSECT** ir **EXCEPT**

**Dekarto sandauga** $l_1(R_1) \times l_2(R_2)$ yra ypatingas junginio atvejis:

**SELECT** $R_1$, $R_2$ **FROM** $l_1$, $l_2$

Galima pradžioje apibrėžti Dekarto sandaugą SQL kalba. Tada lentelių junginys SQL išreiškiamas per išrinkimo, projekcijos ir Dekarto sandaugą.