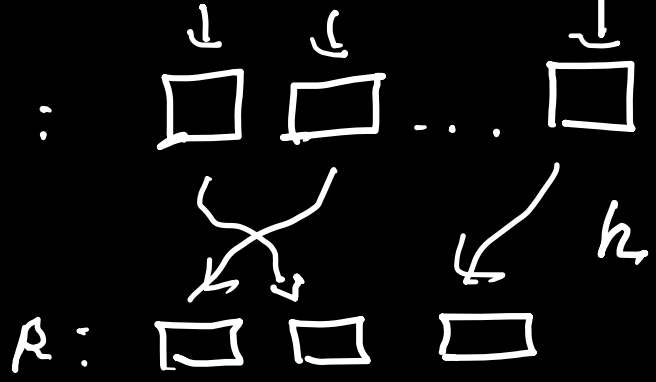
# Przykład (inverted index)

$d_1, d_2, \ldots, d_n$ – dokumenty

word → [S] → lista dok. zaw. słowo w




```
map (id, doc) {
    for w in doc do
        emit (w, id)
}
```

```
reduce (key, L) {
    emit (key, L)
}
```

R $[(w_{i_1}, id_1), \ldots (w_{i_N}, id_N)]$

$[(w_1, L_1), (w_2, L_2), \ldots]$

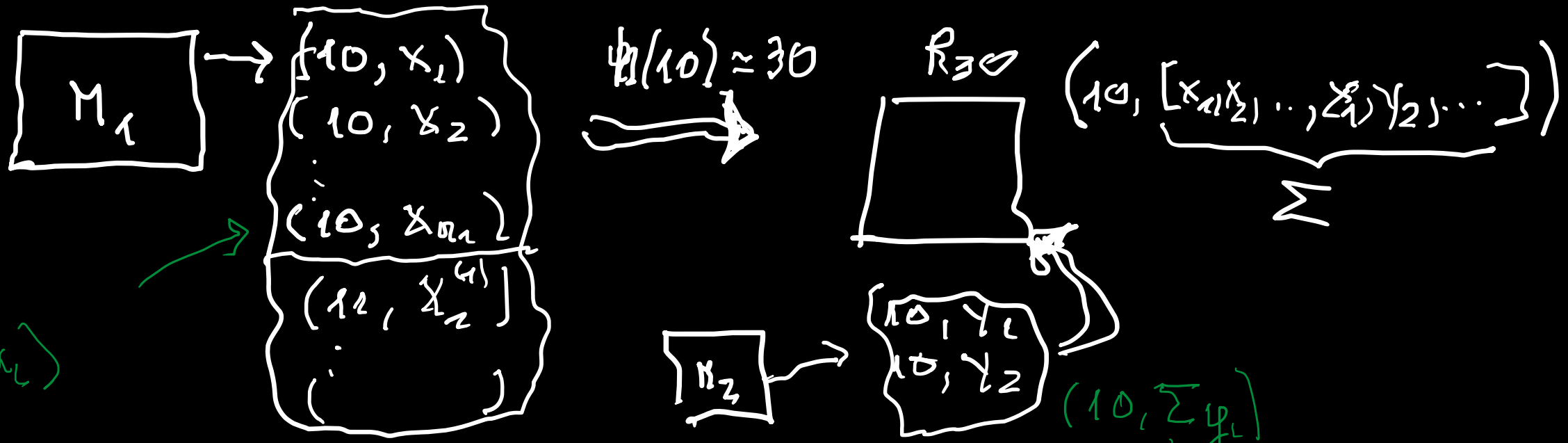
P. Aggregaty : dane $[x_1, x_2] \dots , x_N]$ – ciąg doubli

- suma
- max
- min

$\varphi(x)$ = "ostatnie 10 bitów"

```
map (x) { emit ( φ(x), x); }
reducer (K, L) { emit (k, ΣL, min L, max L); }
```

$L = [x_{L_1}, x_{L_2}, \dots, x_{L_q}]$



$\sum L = \sum x_i + \sum y_i + \sum z_i$ łączność +

COMPOSERS: wad po stronie mapperów po obrobieniu danych i przed rozesłaniem.

```
reducer (k, L) {
    emit (K, Θ(L));
}
```

$$\begin{cases} \Theta(x_1, x_2, \ldots, x_4) = \\ = \Theta\left[c(x_1, \ldots, x_{i_1}), c(x_{i+1}, \ldots x_{i_2}), \ldots\right) \end{cases}$$

$\Theta \Leftarrow$ łączne, przemienne

SUMA

composer (k, L) { emit (k, ΣL); }

MIN : Min (x_1, ..., x_n) y_1 ... y_n) =

= min (min (x), min (y))

MAX : Tą czna

---

Chcemy wyznaczyć Σ, Min, Max

map (x) { emit (φ(x), (x, x, x)); }

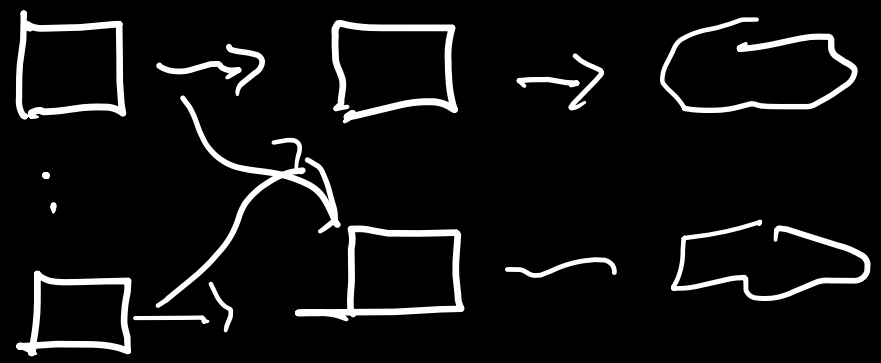
comp (k, L) { emit (k, Σ π_1(L), min π_2(L),

max (π_3(L))}

reduce (k, L) { ... }

L = [(x_1, x_1, x_1), (x_2, x_2, x_2), ...]

ŚREDNIA : $\frac{x_1 + \ldots + x_n}{n} = s(\bar{x})$



$s(\text{całe})$

średnia nie jest łączna !

$s([s(x), s(y)]) \neq s(x \| y)$

map (x) { emit (φ(x), (1, x) }

reduce (K, L)

~~{ emit (k, (length(L), Σ x_L) ); }~~

$L = [(n_1, x_1), (n_2, x_2) \ldots (n_k, x_k)]$

{ emit (k, $\sum \pi_1(L)$, $\sum \pi_2(L)$ ); }

comp (K, L) { emit (k, (length(L), ΣL) ); }

WARIANCJA:

$$|\bar{x}| = n$$

$$Var(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \underbrace{\frac{x_1 + \ldots + x_n}{n}}_{\mu}\right)^2 =$$

$$= \frac{1}{n}\left(\sum_i x_i^2 - 2\sum_i x_i\mu + \sum_i \mu^2\right) = \frac{1}{n}\sum_i x_i^2 - 2\mu\cdot\mu^2 + \mu^2 =$$

$$= \frac{1}{n}\sum x_i^2 - \mu^2 = \frac{1}{n}\sum x_i^2 - \left(\frac{\sum x_i}{n}\right)^2.$$

$$Var(X) = E(X^2) - (E(X))^2$$

```
map(x){
  emit(φ(x), {1, x, x^2});
}
```

Q Jak policzyć wariancję?

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^3 \quad ?$$

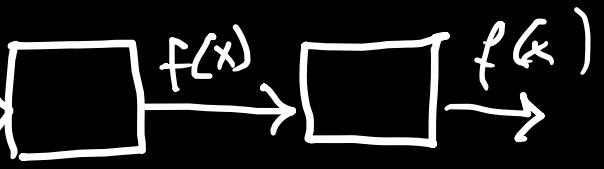
$$x \to (\varphi(x), 1, x, x^2, x^3)$$

---

CO MOŻEMY ZROBIĆ W M-R ?

Mamy zbiory $A, B$.

Operacje mnogościowe: $A \cup B$, $A \cap B$, $A \setminus B$




INPUT: $a \in A$: $(\text{"}A\text{"}, a)$

$b \in B$: $(\text{"}B\text{"}, b)$

```
map (Z, x) {
    emit (x, Z);
}
```

suma : reducer (C, L) { emit (C, 1); }  $A \cup B$

L = [A]
L = [B]
L = [A, B], L = [B, A]

iloczyn $A \cap B$ reduce (C, L) { if |L| = 2 then emit (C, 1); }

różn $A \setminus B$ —— " —— { if L = [A] — " — ; }

Selekcija:

$R(a_1 \dots a_n)$

SELECT $a_1 \dots a_n$ FROM R where $\theta(a_1 \dots a_n)$

map $(a_1 \dots a_n)$ { if $\theta(a_1 \dots a_n)$ emit $(\varphi(\bar{a}), \bar{a})$; }

JOINS

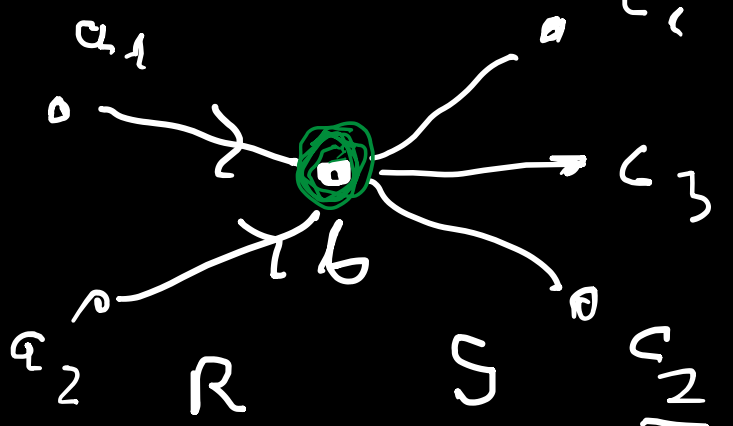
$R(a, b)$, $S(b, c)$

```
{ select R.a, S.c FROM R, S
{ where  R.b = S.b ;
```

map : $(R, a, b) \rightarrow (b, [R, a])$

$(S, b, c) \rightarrow (b, (S, c))$

```
map(X, x, y){
  if (X == "R") {emit(y, ("R", x));}
  else          {emit(x, ("S", y));}
}
```




reducer b : $[(R, a_1), (S, c_1), (R, a_2), (R, a_3), (S, c_2), \ldots]$

reduce(b, L) { 1. $L_1 = sort(L, 1)$

2. $L_1 \longrightarrow L_{11} \| L_{12}$

3. for all $(R, a)$ in $L_{11}$ forall $(S, c)$ in $L_{12}$ : emit $(1, (a, c))$; }

$[(R, a_1), (R, a_2), \ldots (R, a_n)]$

$[(S, c_1), \ldots (S, c_m)]$

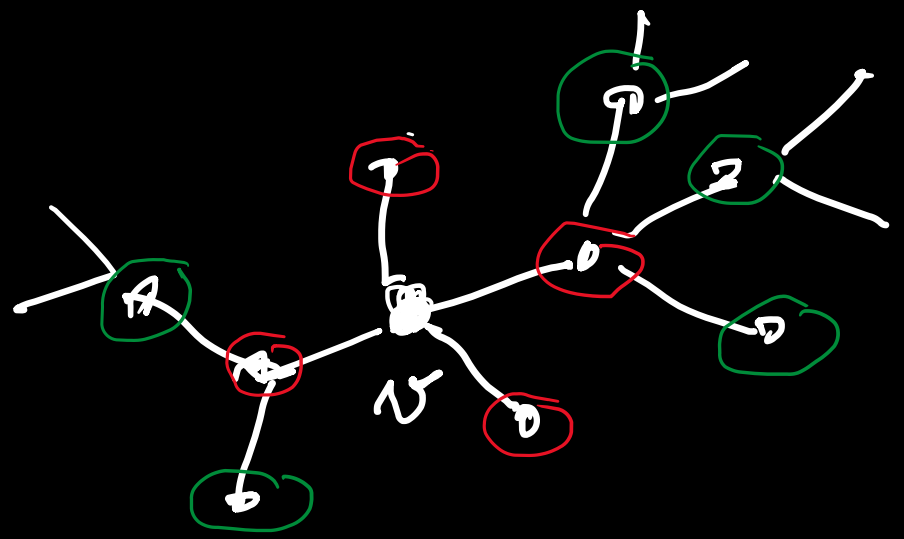
Typowe zastosowanie:

$G = (V, E)$ ← graf.

Dane: $(v_1, v_2)$ ←

22

$(v_1, v_2)$ , $(v_2, v_1)$



$E^* = E \cup flip(E)$ : Relacja

Join : dla $R = E^*$ i $S = E^*$ Zadanie

napisz to dobrze.