

reduce set

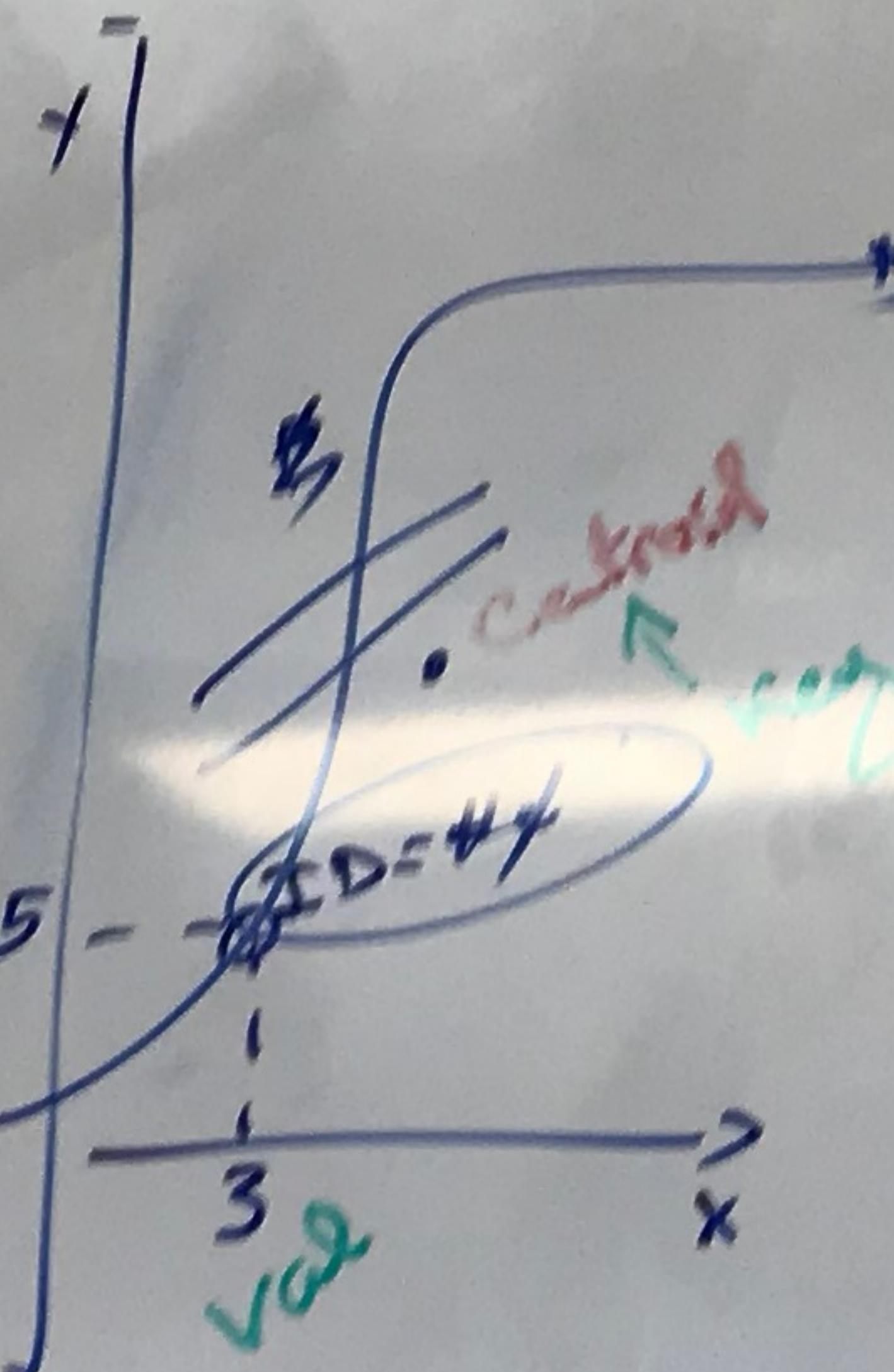
map  $((URL_i, s_i)) : h_i = \text{extract\_host}(URL_i)$

emit  $((h_i, s_i))$

/p/t/r.html → host<sub>1</sub>

/d/r.png, 20 → host<sub>2</sub>

host<sub>2</sub>.com/p<sub>1</sub>/t.r.html, 10  
host<sub>2</sub>.com/p<sub>2</sub>/r.png, 10  
twit.com/



reduce  $((h_i, [s_1, s_2, \dots, s_j])) : \Sigma$   
for  $f-s = \frac{\text{size}}{\text{size}}$  in Ter(3) :  
 $f-s + = \text{size}$   
emit  $((h_i, f-s))$

$\{\bar{P}_1, \bar{P}_2, \dots, \bar{P}_n\}$

$\{C_1, C_2, \dots, C_n\}$

map  $(key, Val) \Rightarrow \text{emit}((id, name, count, size))$   
(reduce  $(id)$ )  $\{G_1, G_2, G_3\}$

*size about 9*

$(URL_1, s_1, -)$   
 $(URL_2, s_2, -)$

$|$   
 $(URL_N, s_N, -)$

$\left[ \text{map } ((URL_i, s_i)) \right.$   
 $h_i = \text{extract\_host}(URL_i)$

$\left. \text{emit } ((h_i, s_i)) \right]$

*function*

$\text{reduce } ((h_i, [s_1, s_2, \dots, s_j])) : \Sigma$   
 for  $s = \text{size}$  in lex(s) :  
 $f - s + = \text{size}$   
 $\text{emit } ((h_i, f - s))$

*return value*

$P://\overline{\text{host}}/\overline{p}/t/r.html \rightarrow \text{host}_2$

$P://\text{host}/d/r.pug, 20 \rightarrow \text{host}_2$

$\uparrow$

$\text{insta.com}/p1/t.pug, 10$   
 $\text{insta.com}/p2/r.pug, 20$   
 $\text{twit.com}/$

$(h_1, s_1)$   
 $(h_2, s_2) \rightarrow (h_2, (s_1 + s_2))$   
 $(h_1, s_3)$