

Dictionary Compression

① CREATE DOMAIN

② Dictionaries

Ⓐ Schema

Ⓑ physical level

Colleagues		
<u>name</u>	street	city
peter	narrowstreet	new york
steve	macstreet	cupertino
mike	longstreet	saarbruecken
tim	unistreet	saarbruecken
hans	msstreet	new york
jens	meerweinstreet	cupertino
frank	narrowstreet	new york
olaf	macstreet	saarbruecken
stefan	unistreet	saarbruecken
alekh	unistreet	saarbruecken
felix	macstreet	new york
jorge	narrowstreet	saarbruecken

Dictionary Compression



Colleagues2		
<u>name</u>	street	cityID
peter	narrowstreet	0
steve	macstreet	1
mike	longstreet	2
tim	unistreet	2
hans	msstreet	0
jens	meerweinstreet	1
frank	narrowstreet	0
olaf	macstreet	2
stefan	unistreet	2
alekh	unistreet	2
felix	macstreet	0
jorge	narrowstreet	2

Cities_Dictionary	
<u>cityID</u>	city
0	new york
1	cuppertino
2	saarbruecken

Dictionary Compression

CREATE VIEW

Colleagues3		
<u>name</u>	streetID	cityID
peter	0	0
steve	1	1
mike	2	2
tim	3	2
hans	4	0
jens	5	1
frank	0	0
olaf	1	2
stefan	3	2
alekh	3	2
felix	1	0
jorge	0	2



Cities_Dictionary	
<u>cityID</u>	city
0	new york
1	cupertino
2	saarbruecken

Streets_Dictionary	
<u>streetID</u>	streets
0	narrowstreet
1	macstreet
2	longstreet
3	unistreet
4	msstreet
5	meerweinstreet

Example Query Rewrite

SELECT
FROM
WHERE

name
Colleagues
street = 'unistreet';

becomes:

SELECT
FROM
WHERE

name
Colleagues3
streetID =
(**SELECT**
FROM
WHERE
);

streetID
Streets_Dictionary
streets='unistreet'

Colleagues		
name	street	city
peter	narrowstreet	new york
steve	macstreet	cupertino
mike	longstreet	saarbruecken
tim	unistreet	saarbruecken
hans	msstreet	new york
jens	meerweinstre	cupertino
frank	narrowstreet	new york
olaf	macstreet	saarbruecken
stefan	unistreet	saarbruecken
alekh	unistreet	saarbruecken
felix	macstreet	new york
jorge	narrowstreet	saarbruecken

Colleagues3		
name	streetID	cityID
peter	0	0
steve	1	1
mike	2	2
tim	3	2
hans	4	0
jens	5	1
frank	0	0
olaf	1	2
stefan	3	2
alekh	3	2
felix	1	0
jorge	0	2

Streets_Dictionary	
streetID	streets
0	narrowstreet
1	macstreet
2	longstreet
3	unistreet
4	msstreet
5	meerweinstreet

Example Query Rewrite

SELECT
FROM
GROUP BY

city, count(*)
Colleagues
city;

0, 4
1, 2
2, 6

becomes:

WITH

T AS (

SELECT cityid, count(*) AS c
FROM Colleagues3
GROUP BY cityid;

)
SELECT
FROM

city, c
T NATURAL JOIN Cities_Dictionary;

Colleagues		
name	street	city
peter	narrowstreet	new york
steve	macstreet	cupertino
mike	longstreet	saarbruecken
tim	unistreet	saarbruecken
hans	msstreet	new york
jens	meerweinstre	cupertino
frank	narrowstreet	new york
olaf	macstreet	saarbruecken
stefan	unistreet	saarbruecken
alekh	unistreet	saarbruecken
felix	macstreet	new york
jorge	narrowstreet	saarbruecken

Colleagues3		
name	streetID	cityID
peter	0	0
steve	1	1
mike	2	2
tim	3	2
hans	4	0
jens	5	1
frank	0	0
olaf	1	2
stefan	3	2
alekh	3	2
felix	1	0
jorge	0	2

Cities Dictionary	
cityID	city
0	new york
1	cupertino
2	saarbruecken

Domain Encoding

int

$$2^2 = 4$$

Colleagues3		
<u>name</u>	streetID	cityID
peter	0	0
steve	1	1
mike	2	2
tim	3	2
hans	4	0
jens	5	1
frank	0	0
olaf	1	2
stefan	3	2
alekh	3	2
felix	1	0
jorge	0	2

CREATE DOMAIN

3 bits

$$2^3 = 8$$

Dictionary Compression

Advantages:

- string data converted to numeric data

- single row access still possible

- additional compression methods may be applied

- can be exploited for query processing

Disadvantages:

- savings depend on #duplicates/#entries

- extra joins to dictionary