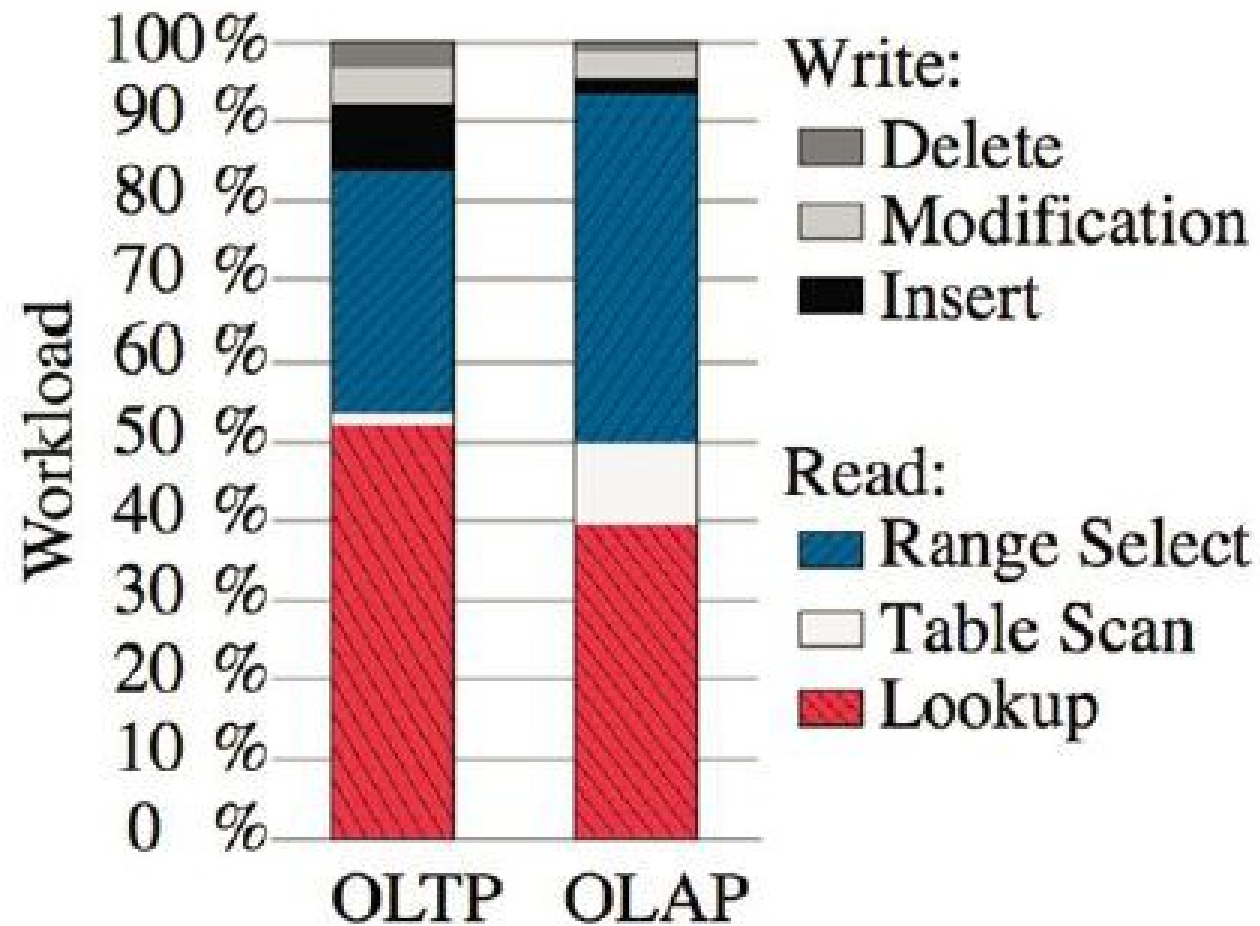


IN-MEMORY DATAMANAGEMENT

Úvod – Pavel HAJN

OLAP a OLTP



Column based storage – uložení dat

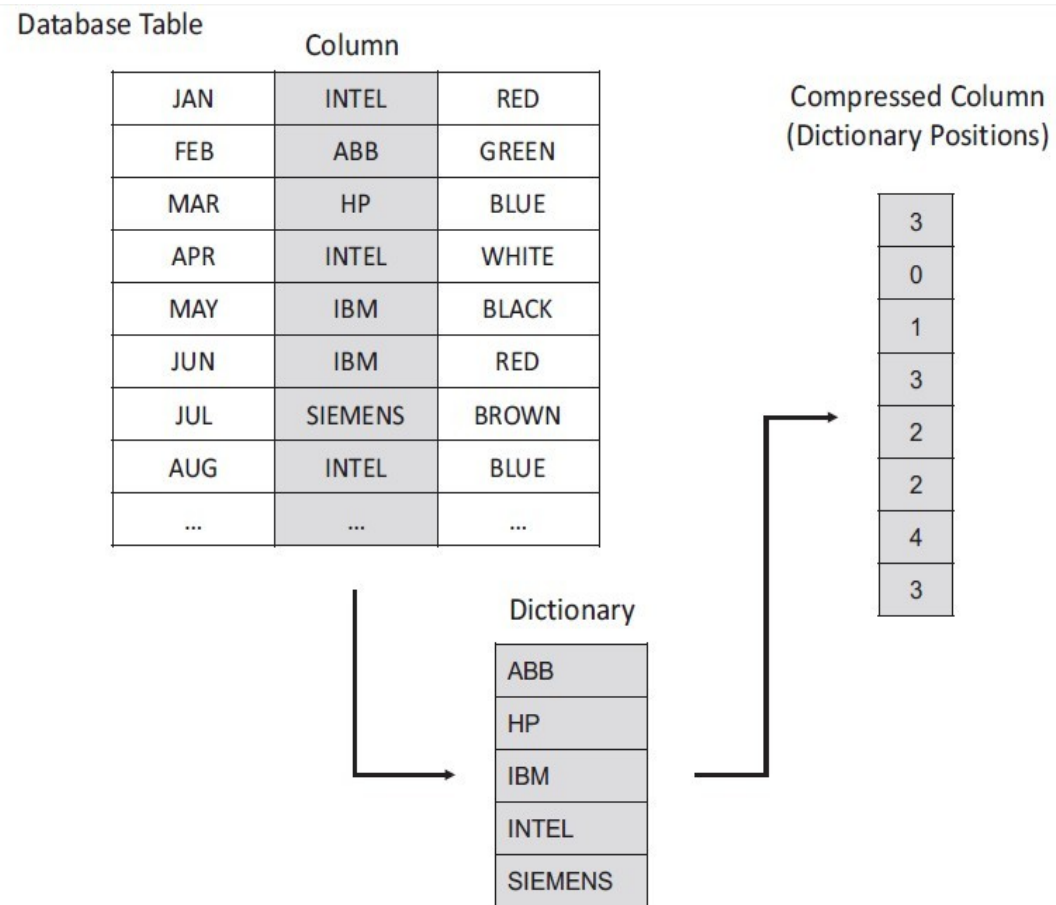


Figure 4.16: Dictionary Encoding

Příklad pro světovou populaci

Column	Cardinality	Bits Needed	Item Size	Plain Size	Size with Dictionary (Dictionary + Column)	Compression Factor
First names	5 millions	23 bit	50 Byte	400GB	250MB + 23GB	≈ 17
Last names	8 millions	23 bit	50 Byte	400GB	400MB + 23GB	≈ 17
Gender	2	1 bit	1 Byte	8GB	2b + 1GB	≈ 8
City	1 million	20 bit	50 Byte	400GB	50MB + 20GB	≈ 20
Country	200	8 bit	47 Byte	376GB	9.4kB + 8GB	≈ 47
Birthday	40000	16 bit	2 Byte	16GB	80kB + 16GB	≈ 1
Totals			200 Byte	≈ 1.6TB	≈ 92GB	≈ 17

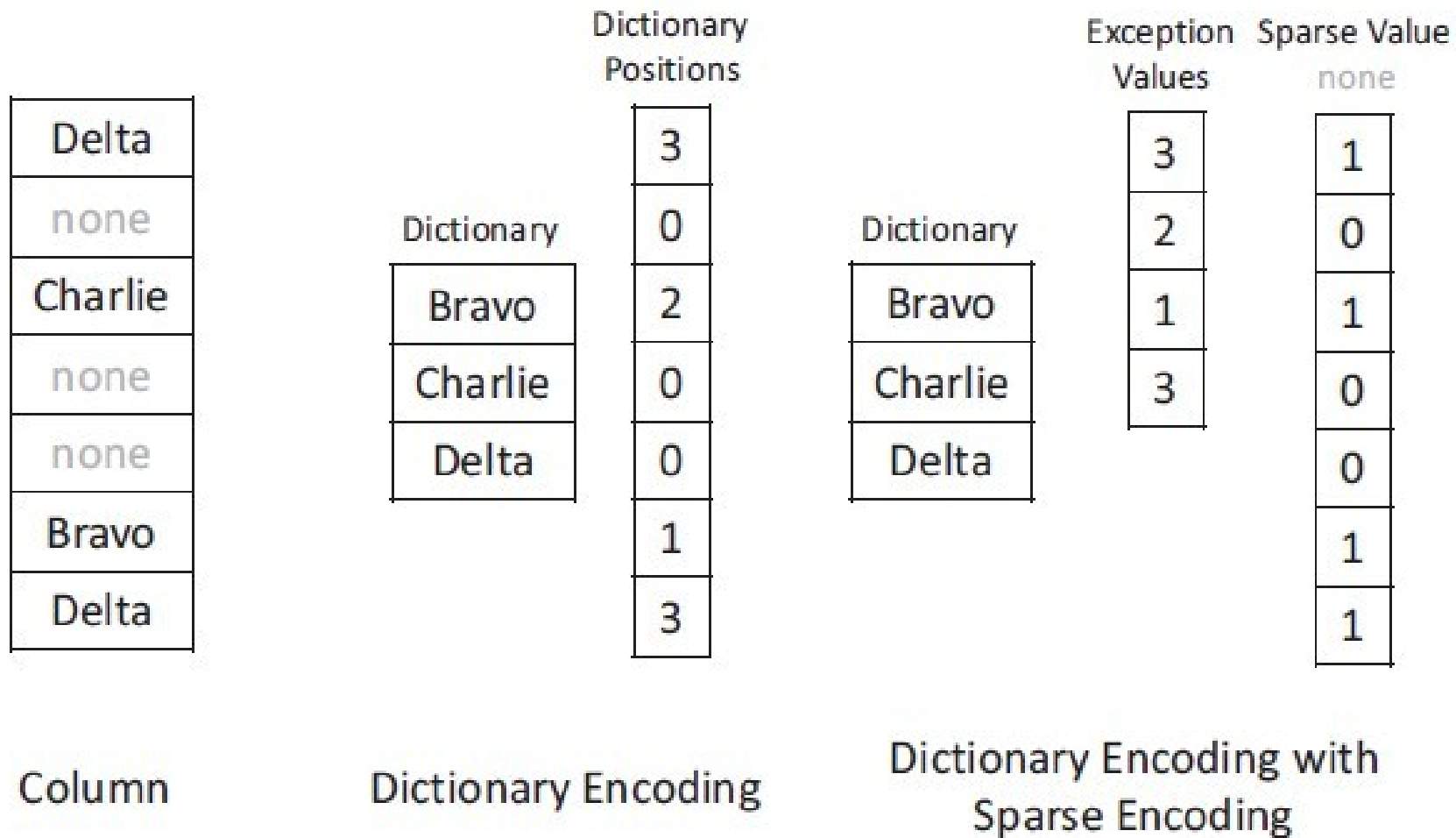


Figure 4.17: Sparse Encoding

Exekuční plán dotazu

Value ID	Dictionary for "country"
0	Algeria
1	France
2	Germany
3	Italy
4	Netherlands
...	...

id	fname	lname	country	gender
2394	Gianluigi	Buffon	3	1
3010	Lena	Gercke	2	0
3040	Mario	Balotelli	3	1
3949	Manuel	Neuer	2	1
4902	Lukas	Podolski	2	1
20102	Klaas-Jan	Huntelaar	4	1

Value ID	Dictionary for "gender"
0	f
1	m

gender = 1 ("m")

country = 3 ("Italy")

Position
0
2

AND

Position
0
2
3
4
5

Position
0
2



Agregační funkce

SELECT country, COUNT(*) AS citizens FROM world_population
GROUP BY country

recID	fname	lname	gender	city	country	birthday
0	John	Smith	m	Chicago	USA	12.03.1964
1	Mary	Brown	f	London	UK	12.05.1964
2	Jane	Doe	f	Palo Alto	USA	23.04.1976
3	John	Doe	m	Palo Alto	USA	17.06.1952
4	Peter	Schmidt	m	Potsdam	GER	11.11.1975
...

Attribute Vector for "country"

44	43	44	44	42	...
----	----	----	----	----	-----

Result Map

44	3
43	1
42	1
...	...

+

country Dictionary

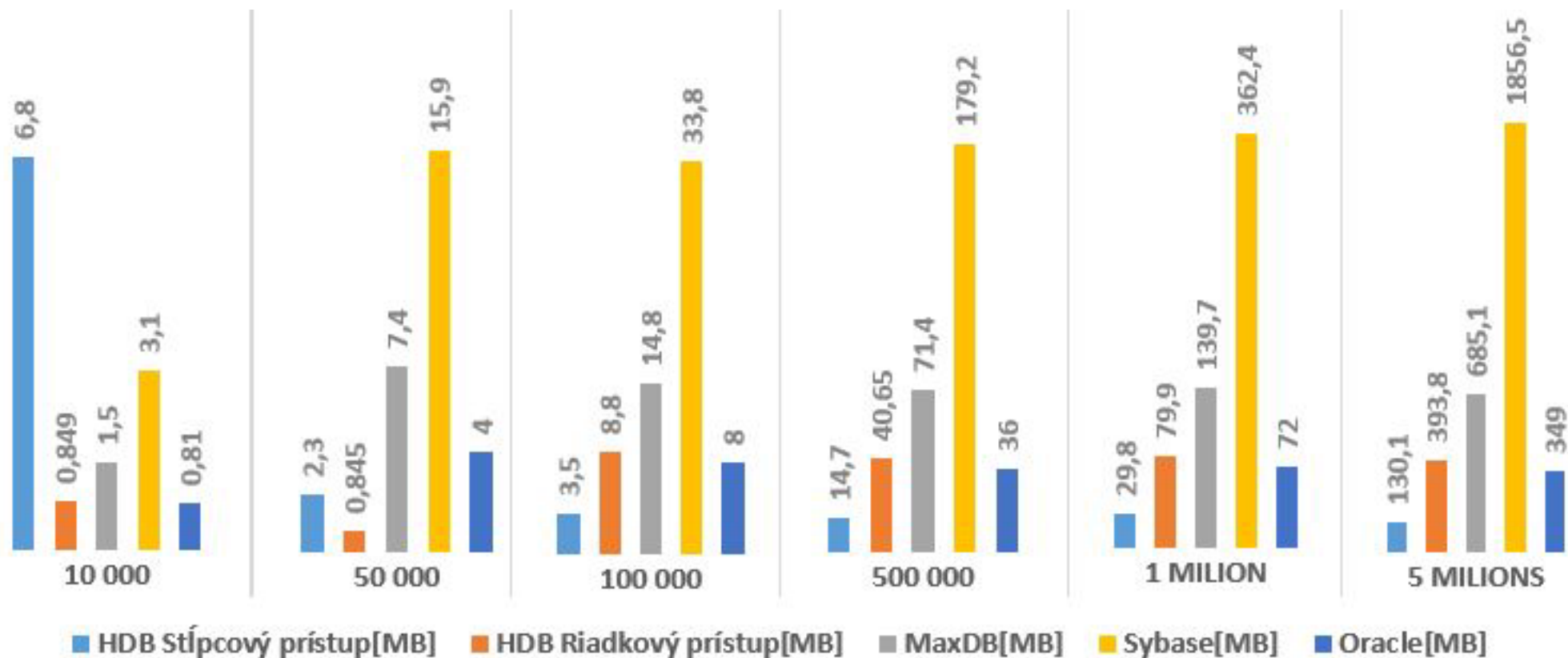
valueID	value
...	...
42	GER
43	UK
44	USA
...	...

=

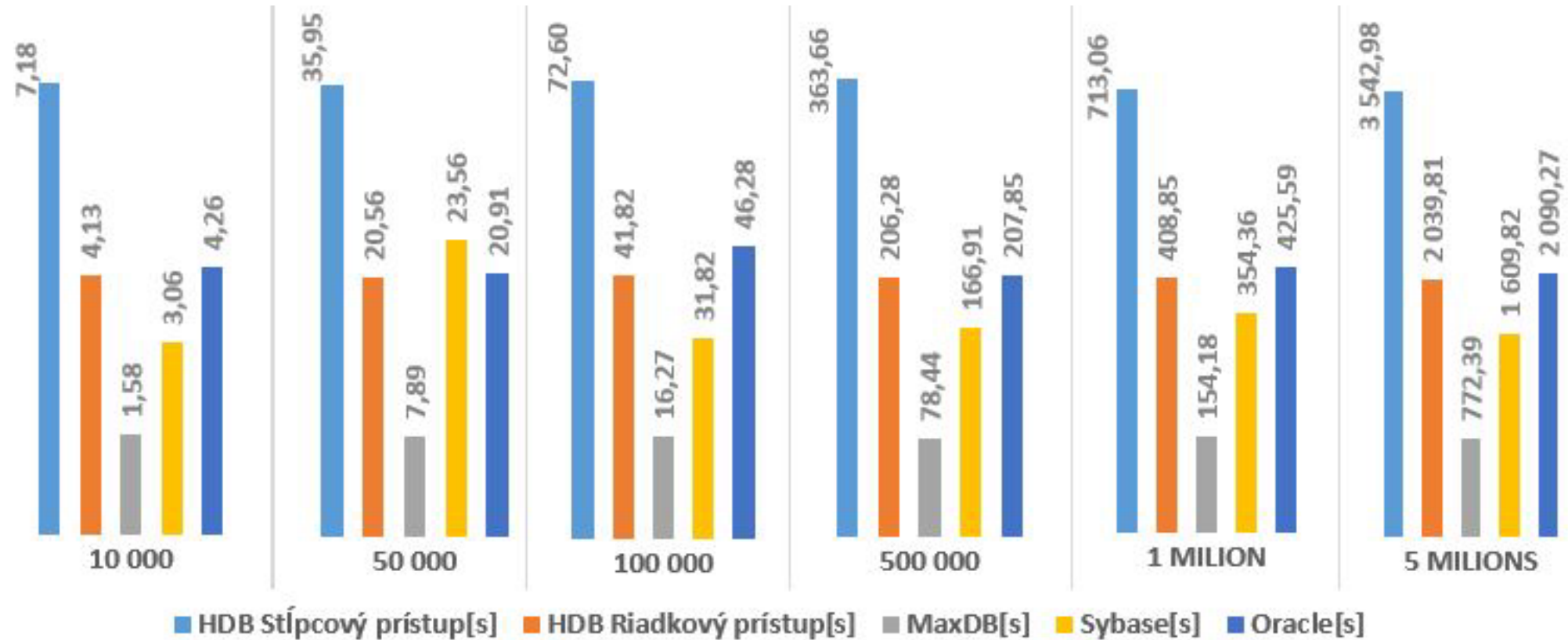
Result

country	citizens
USA	3
UK	1
GER	1
...	...

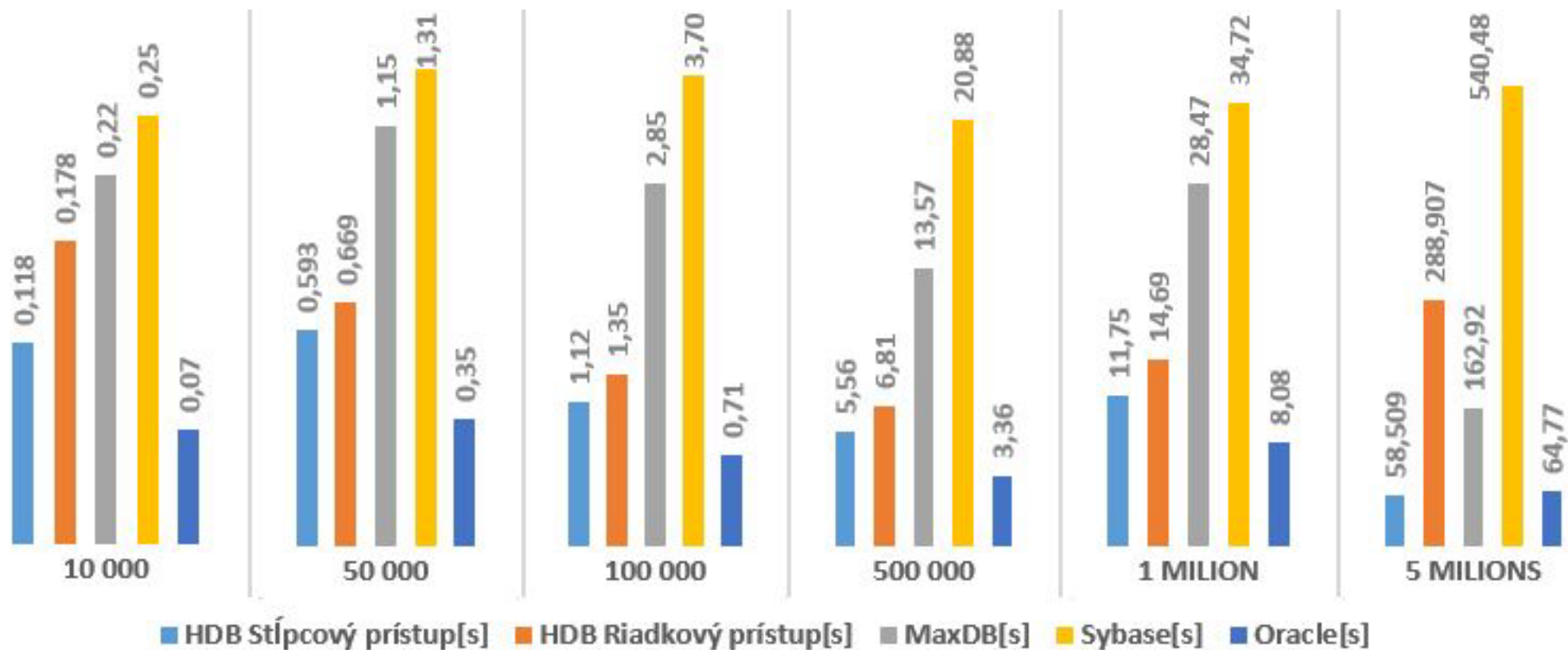
Výkonové testy – velikost DB



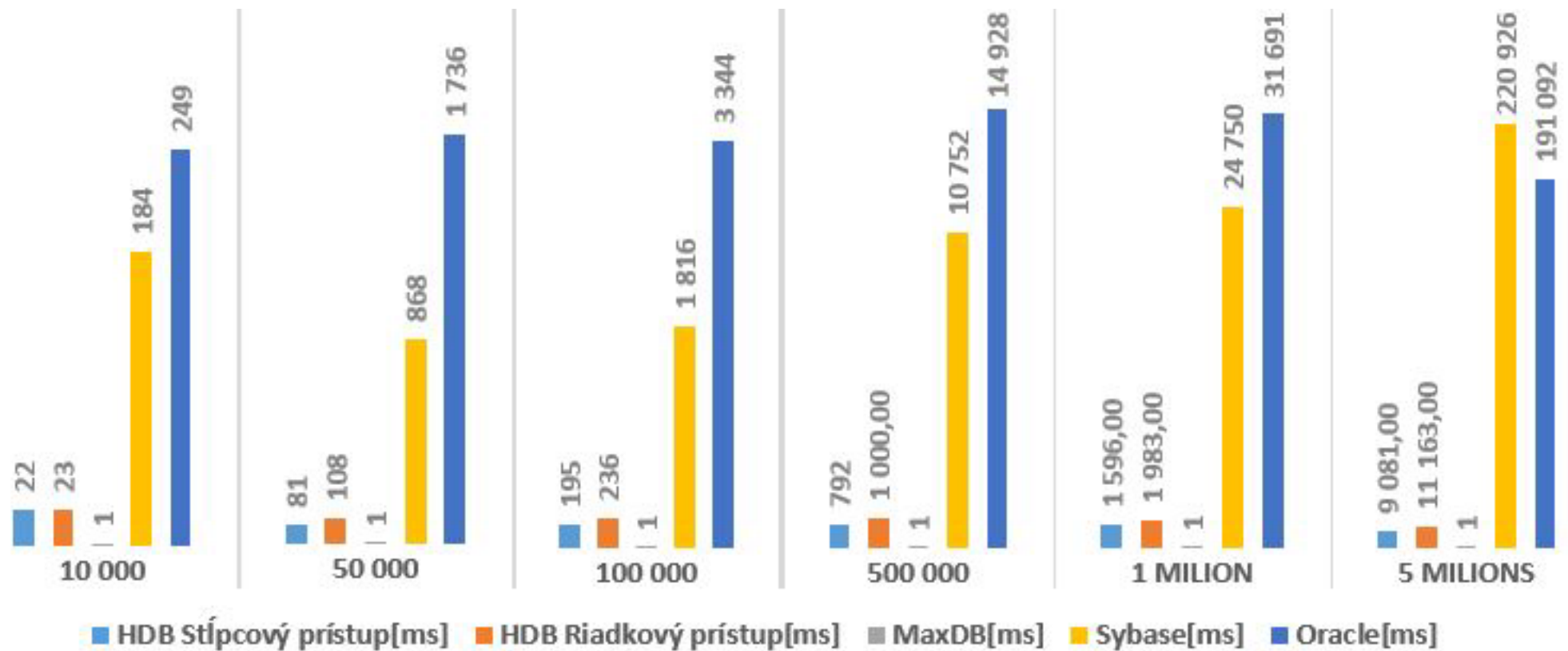
Výkonové testy – INSERT – opakovaný po 1 záznamu



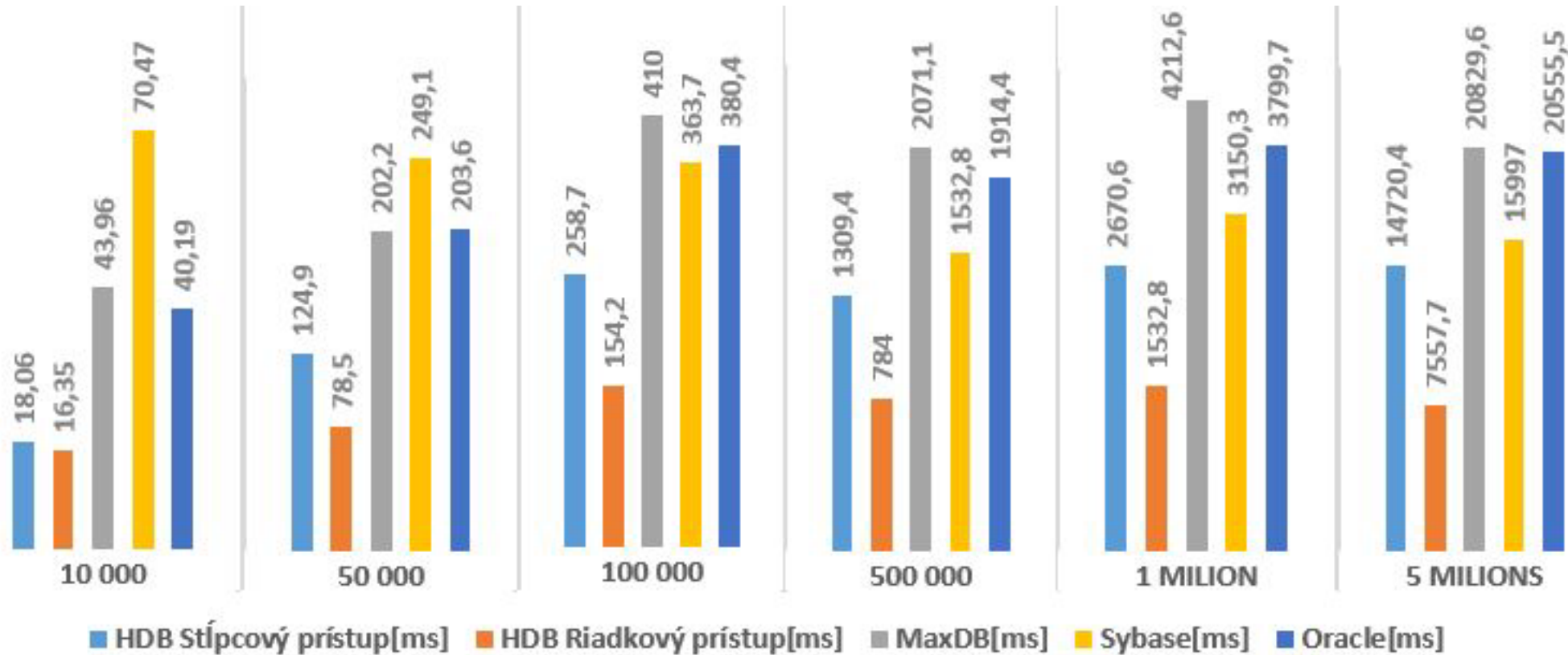
Výkonové testy – INSERT – hromadný



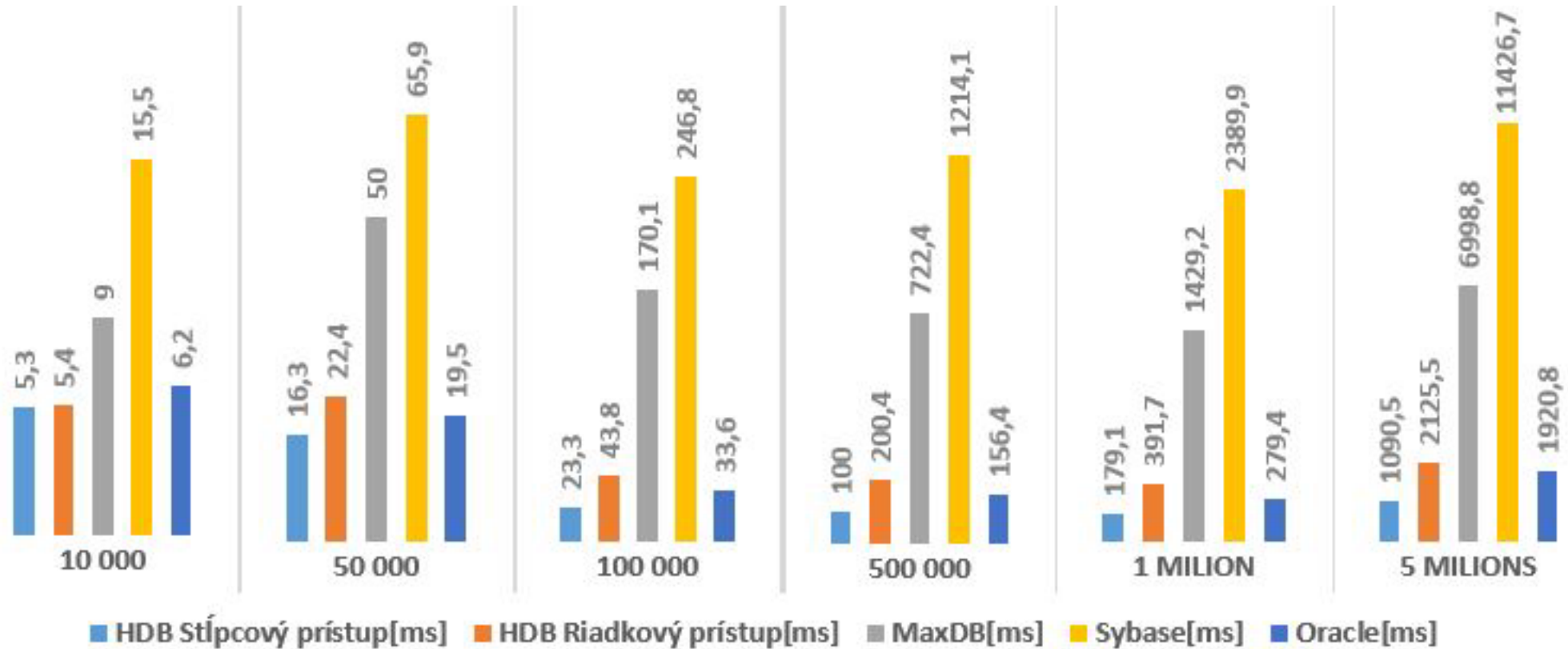
Výkonové testy – DELETE



Výkonové testy – SELECT jednoduchý bez podmínek



Výkonové testy – SELECT složitý – join a výběry



Výkonové testy – SELECT agregační funkce

