A Comparison of Knives for Bread Slicing

Alekh Jindal*, Endre Palatinus, Vladimir Pavlov, Jens Dittrich

Information Systems Group
Saarland University

*MIT CSAIL

Data Layout

Row-layout

CUSTKEY	NATIONKEY	NAME	MKTSEGMENT	ADDRESS	COMMENT	PHONE	ACCTBAL
1234556	DE	G.W.	43	E1.4	-	1234	€987,513
2334444	GB	I.N.	22	OX13	-	332	€10,522
1123234	US	M.S.	22	CA16	-	1233	€6,452
2323454	DE	J.D.	43	E1.3	CST_LOW	54443	€399
2311555	GB	A.M.	12	WA154	-	23442	€46,523
1231235	NL	T.V.	42	AM3321	-	1123	€180,000

Column-layout

CUSTKEY	NATIONKEY	NAME	MKTSEGMENT	ADDRESS	COMMENT	PHONE	ACCTBAL
1234556	DE	G.W.	43	E1.4	-	1234	€987,513
2334444	GB	I.N.	22	OX13	-	332	€10,522
1123234	US	M.S.	22	CA16	-	1233	€6,452
2323454	DE	J.D.	43	E1.3	CST_LOW	54443	€399
2311555	GB	A.M.	12	WA154	-	23442	€46,523
1231235	NL	T.V.	42	AM3321	-	1123	€180,000

Column-grouping

CUSTKEY	NATIONKEY	NAME	MKTSEGMENT	ADDRESS	COMMENT	PHONE	ACCTBAL
1234556	DE	G.W.	43	E1.4	-	1234	€987,513
2334444	GB	I.N.	22	OX13	-	332	€10,522
1123234	US	M.S.	22	CA16	-	1233	€6,452
2323454	DE	J.D.	43	E1.3	CST_LOW	54443	€399
2311555	GB	A.M.	12	WA154	-	23442	€46,523
1231235	NL	T.V.	42	AM3321	-	1123	€180,000

The Vertical Partitioning Problem

- Given a workload and a cost function
- Provide a complete and disjunct partitioning of the set of attributes of a table

Workload

	CUSTKEY	NAME	ADDRESS	NATIONKEY	PHONE	ACCTBAL	MKTSEGMENT	COMMENT
Q1								
Q2								
Q3								
Q4								
Q5								
Q6								
Q7								
Q8								
Q9								
Q10								
Q11								
Q12								
Q13								
Q14								
Q15								
Q16								
Q17								
Q18								
Q19								
Q20								
Q21								
Q22								

Selectivity?

High selectivity → Indexes

Low selectivity → Vertical partitioning

Vertical Partitioning in Legacy Row-Stores

TPC-H Customer

CUSTKEY	NATIONKEY	NAME	MKTSEGMENT	ADDRESS	COMMENT	PHONE	ACCTBAL

P1

CUSTKEY

P2

NATIONKEY

P3

NAME

P4

MKTSEGMENT

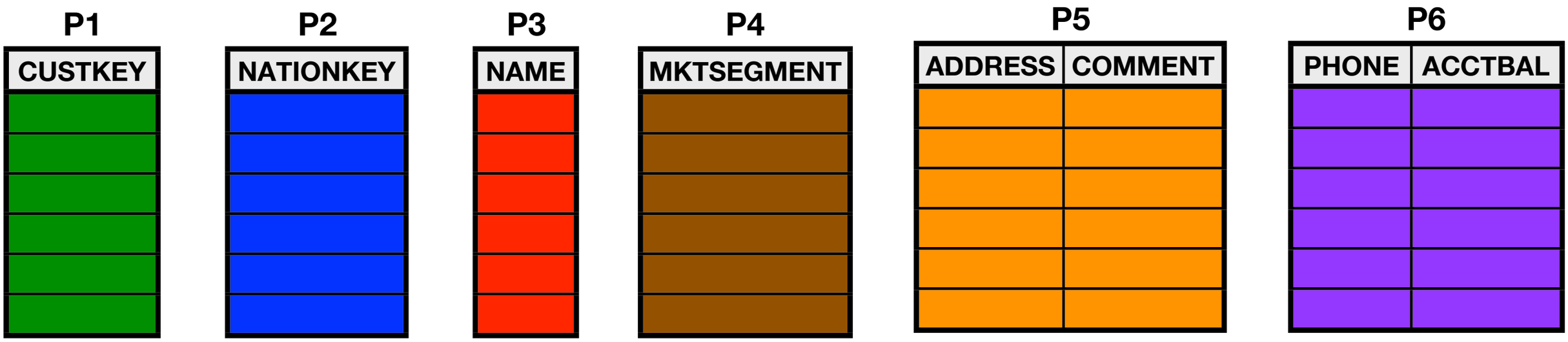
P5

ADDRESS	COMMENT

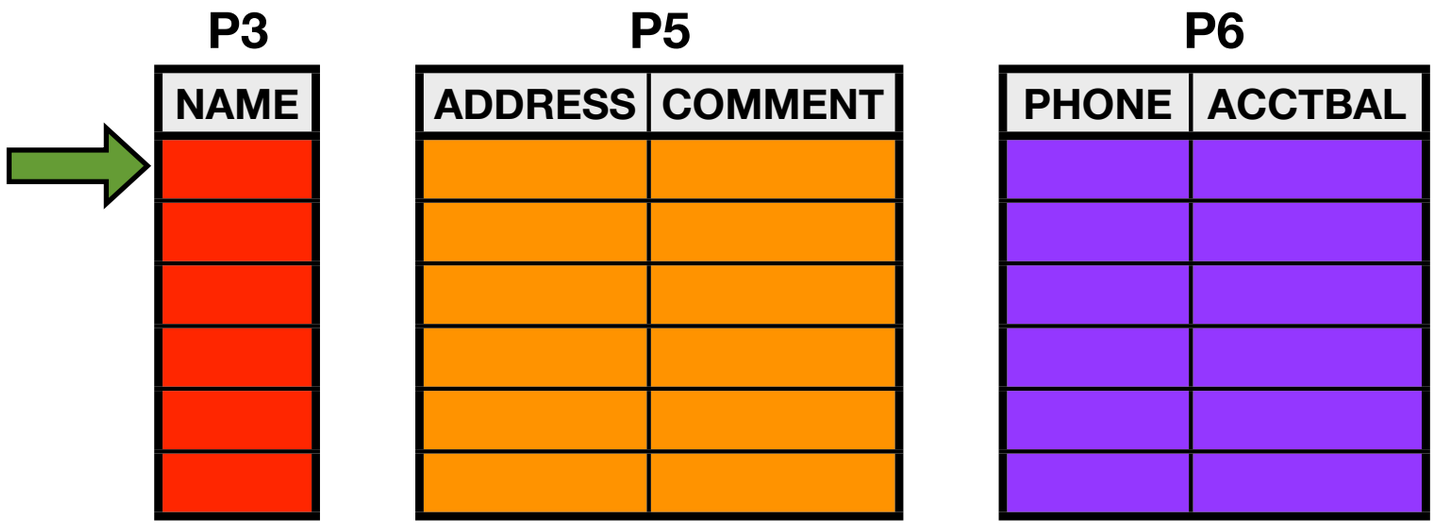
P6

PHONE	ACCTBAL

SELECT Name, Address, Acctbal
FROM Customer



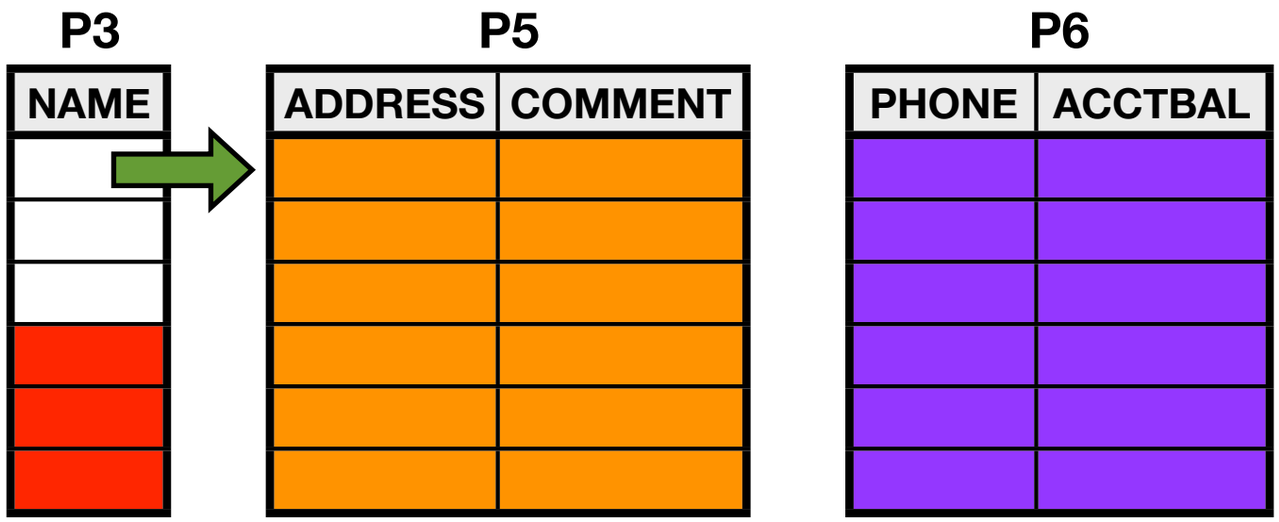
SELECT Name, Address, Acctbal
FROM Customer



Database Buffer

NAME	ADDRESS	COMMENT	PHONE	ACCTBAL

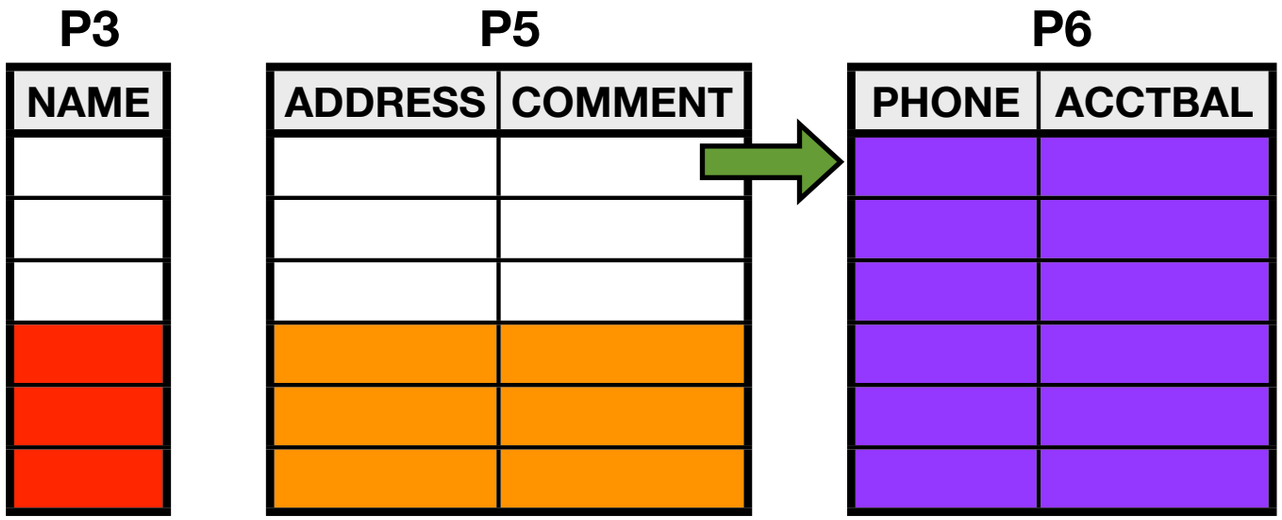
SELECT Name, Address, Acctbal
FROM Customer



Database Buffer

NAME	ADDRESS	COMMENT	PHONE	ACCTBAL

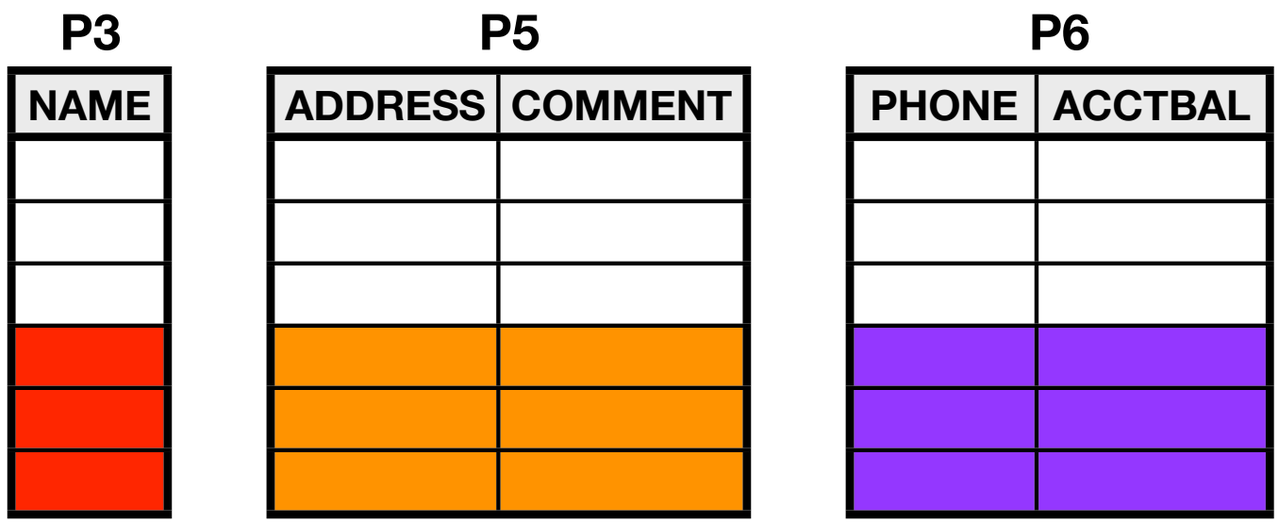
SELECT Name, Address, Acctbal
FROM Customer



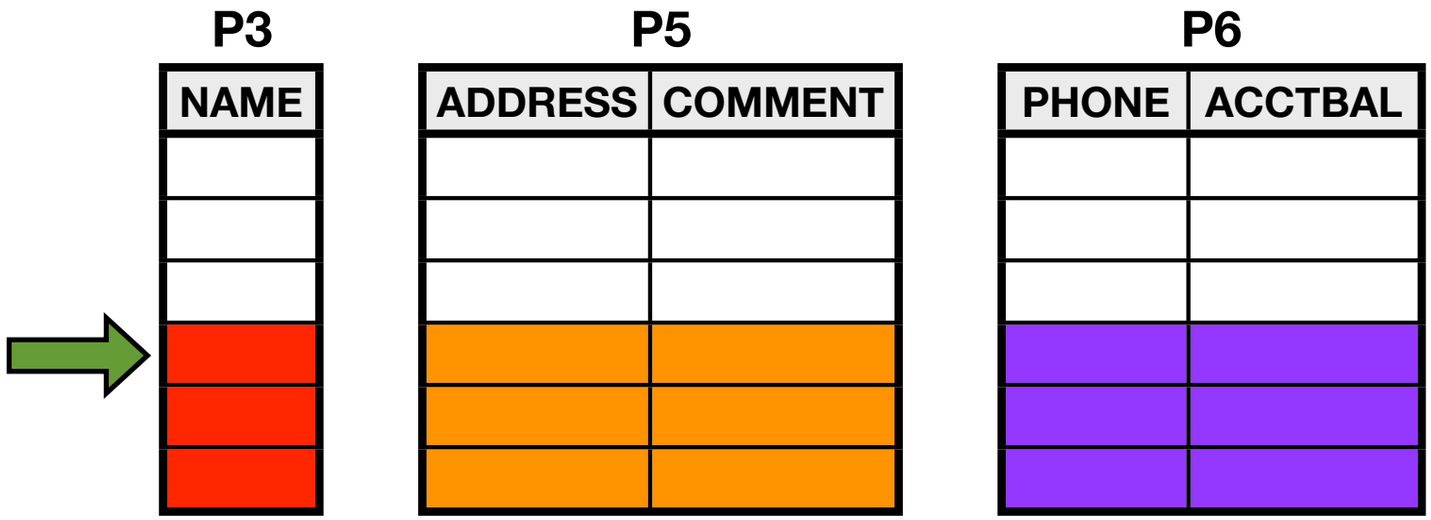
Database Buffer

NAME	ADDRESS	COMMENT	PHONE	ACCTBAL

SELECT Name, Address, Acctbal
FROM Customer



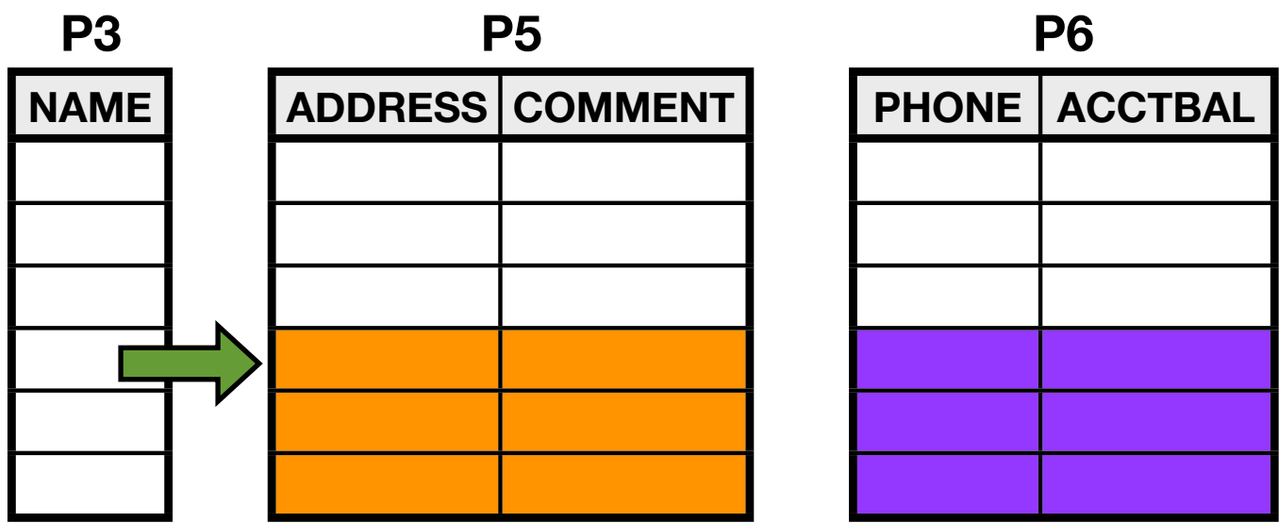
SELECT Name, Address, Acctbal
FROM Customer



Database Buffer

NAME	ADDRESS	COMMENT	PHONE	ACCTBAL

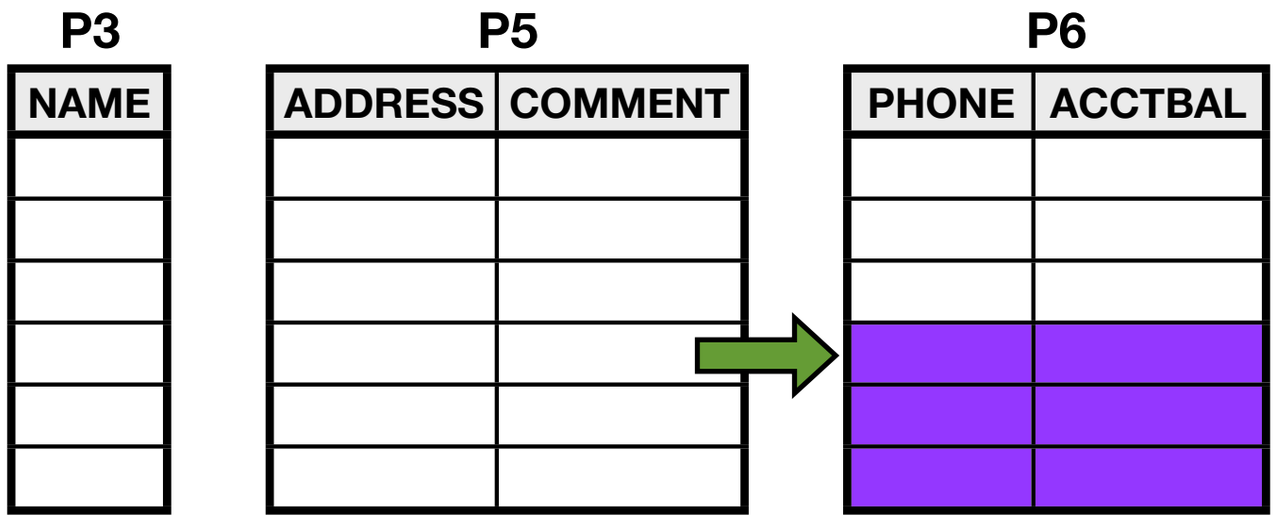
SELECT Name, Address, Acctbal
FROM Customer



Database Buffer

NAME	ADDRESS	COMMENT	PHONE	ACCTBAL

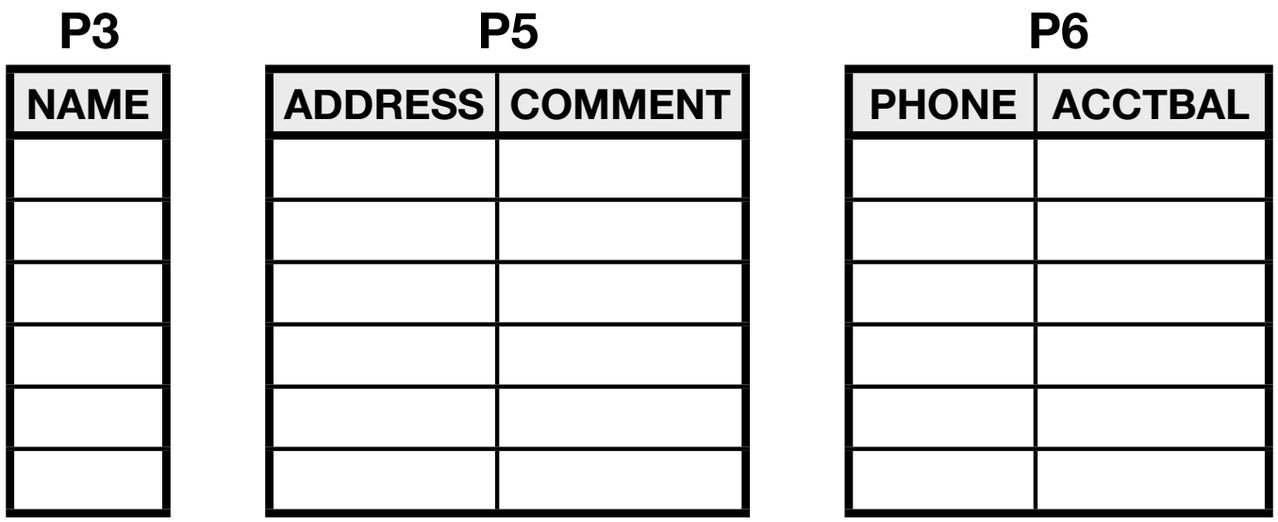
SELECT Name, Address, Acctbal
FROM Customer



Database Buffer

NAME	ADDRESS	COMMENT	PHONE	ACCTBAL

SELECT Name, Address, Acctbal
FROM Customer



Database Buffer

NAME	ADDRESS	COMMENT	PHONE	ACCTBAL

Classification of VP algorithms

Starting Point	Whole workload
	Attribute subset
	Query subset
Search Strategy	Brute force
	Top-down
	Bottom-up
Candidate Pruning	No pruning
	Threshold-based

		AutoPart	HillClimb	HYRISE	Navathe	O2P	Trojan	Brute Force
Starting Point	Whole workload							
	Attribute subset							
	Query subset							
Search Strategy	Brute force							
	Top-down							
	Bottom-up							
Candidate Pruning	No pruning							
	Threshold-based							

HillClimb Example

Iteration 1

	CUSTKEY	NAME	ADDRESS	NATIONKEY	PHONE	ACCTBAL	MKTSEGMENT	COMMENT
Q1								
Q2								
Q3								
Q4								
Q5								
Q6								
Q7								
Q8								
Q9								
Q10								
Q11								
Q12								
Q13								
Q14								
Q15								
Q16								
Q17								
Q18								
Q19								
Q20								
Q21								
Q22								

Iteration 2

	CUSTKEY	NAME	ADDRESS	NATIONKEY	PHONE	ACCTBAL	MKTSEGMENT	COMMENT
Q1								
Q2								
Q3								
Q4								
Q5								
Q6								
Q7								
Q8								
Q9								
Q10								
Q11								
Q12								
Q13								
Q14								
Q15								
Q16								
Q17								
Q18								
Q19								
Q20								
Q21								
Q22								

Iteration 3

	CUSTKEY	NAME	ADDRESS	NATIONKEY	PHONE	ACCTBAL	MKTSEGMENT	COMMENT
Q1								
Q2								
Q3								
Q4								
Q5								
Q6								
Q7								
Q8								
Q9								
Q10								
Q11								
Q12								
Q13								
Q14								
Q15								
Q16								
Q17								
Q18								
Q19								
Q20								
Q21								
Q22								

Settings for VP Algos

Granularity	FILE
	DATA PAGE
	DATABASE BLOCK
Hardware	
Workload	
Replication	
System	

Settings for VP Algos

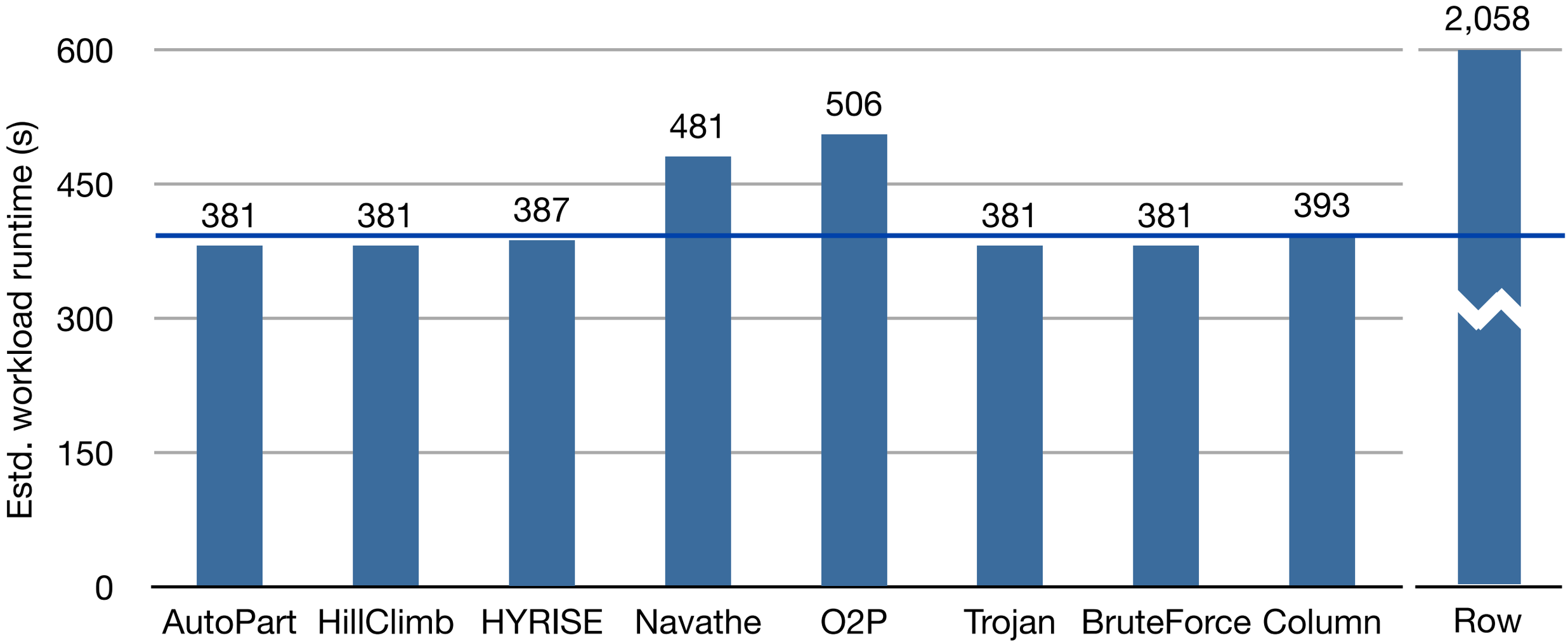
Granularity	FILE
	DATA PAGE
	DATABASE BLOCK
Hardware	HARD DISK
	MAIN MEMORY
Workload	OFFLINE
	ONLINE
Replication	NONE
	FULL
	PARTIAL
System	CUSTOM
	COST MODEL
	OPEN SOURCE

Settings for VP Algos

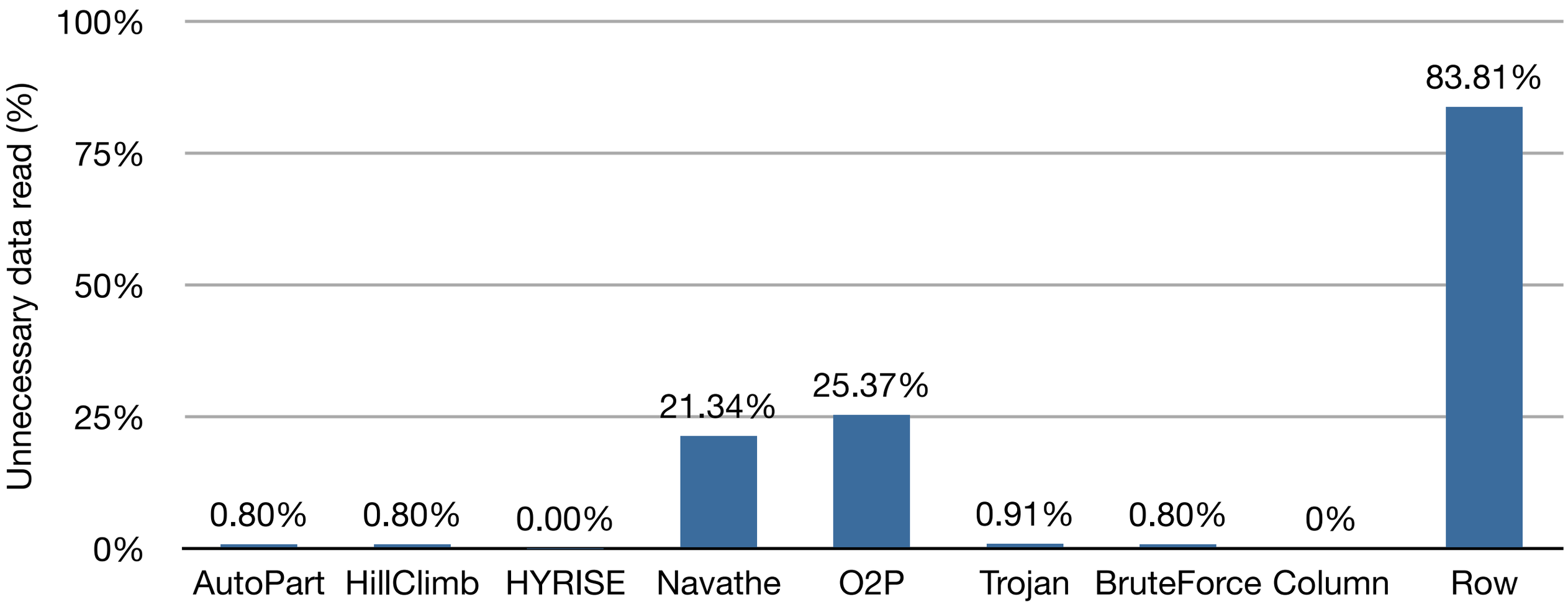
		AutoPart	HillClimb	HYRISE	Navathe	O2P	Trojan
Granularity	FILE	■		■	■	■	
	DATA PAGE		■				
	DATABASE BLOCK						■
Hardware	HARD DISK	■	■		■	■	■
	MAIN MEMORY			■			
Workload	OFFLINE	■	■	■	■		■
	ONLINE					■	
Replication	NONE		■	■	■	■	
	FULL						■
	PARTIAL	■					
System	CUSTOM		■	■			
	COST MODEL	■			■		
	OPEN SOURCE					■	■

Experimental Results

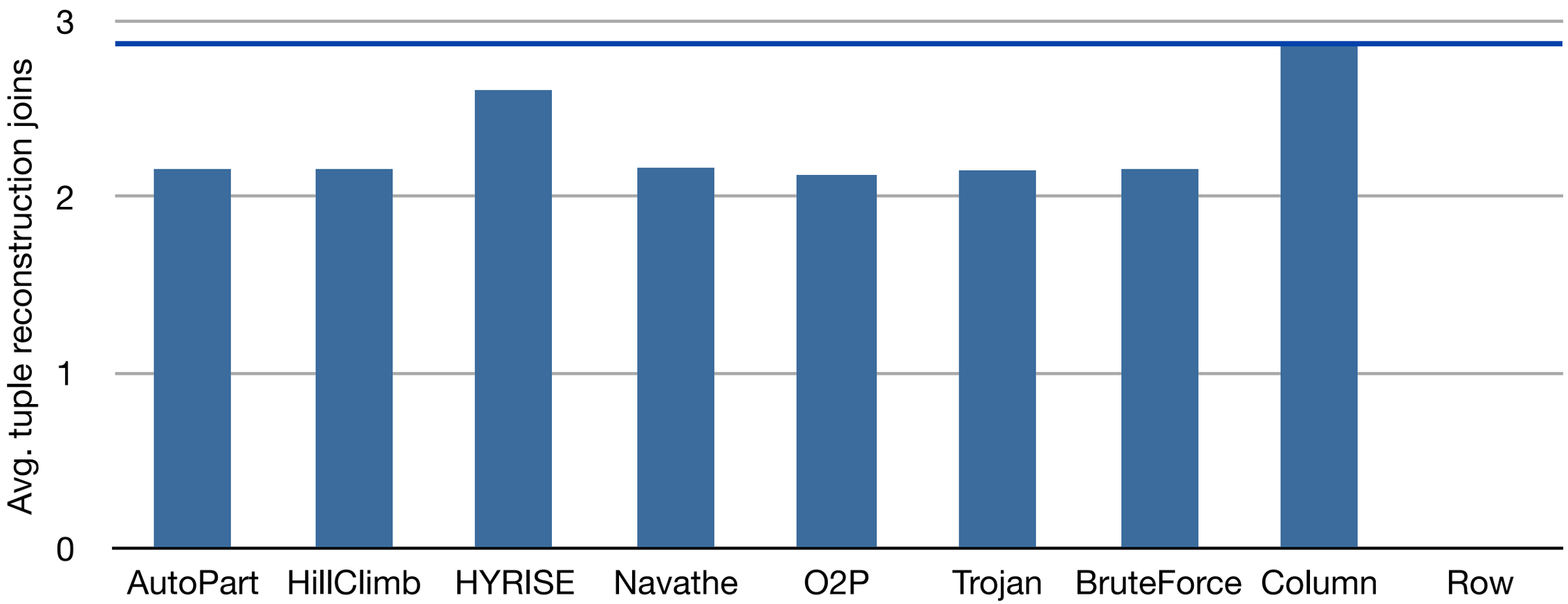
TPC-H Query Runtimes



Unnecessary Data Read



Average #Tuple Reconstruction Joins



Try another Benchmark

Layout	TPC-H	SSB
AutoPart	3.71%	5.29%
HillClimb	3.71%	5.29%
HYRISE	1.58%	5.27%
Navathe	-21.47%	1.64%
O2P	-27.74%	1.64%
Trojan	3.71%	0.05%
BruteForce	3.71%	5.29%

Improvement over Column-layout

[O'Neil et al: Star Schema Benchmark]

Try another Cost Model

Layout	HDD	MM
AutoPart	3.71%	0.00%
HillClimb	3.71%	0.00%
HYRISE	1.58%	0.00%
Navathe	-21.47%	-15.07%
O2P	-27.74%	-15.53%
Trojan	3.71%	0.00%
BruteForce	3.71%	0.00%

Improvement over Column-layout

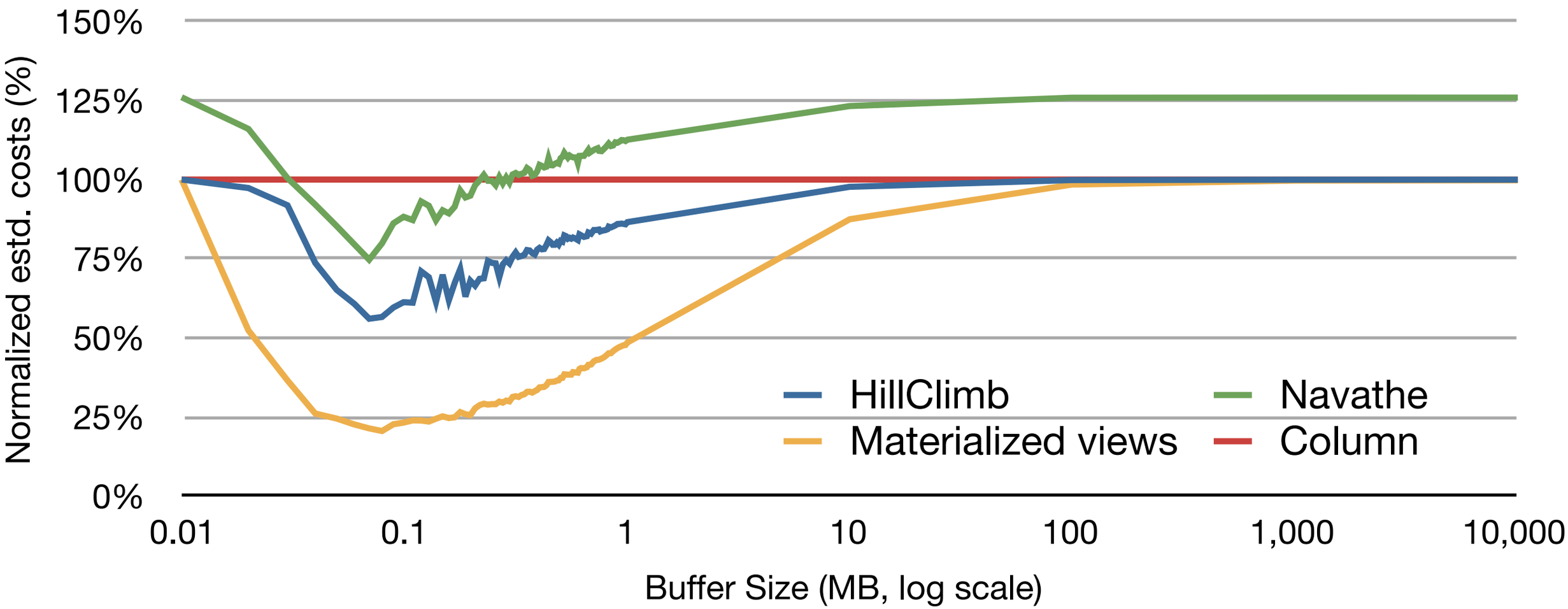
[Grund et al: HYRISE, VLDB'10]

Try it in DBMS-X

Compression	Row	Column	HillClimb
Default (LZO or Delta)	1652 s	377 s	450 s
Dictionary	1265 s	511 s	532 s

Actual Workload Runtimes

Buffer Size is Crucial



Summary

- Buffer size is crucial
- Column layout is good enough
- HillClimb is the best VP algorithm