Automatic Database Partitioning in Parallel OLTP Systems

@andy_pavlo

BROWN

SIGMOD
May 22nd, 2012

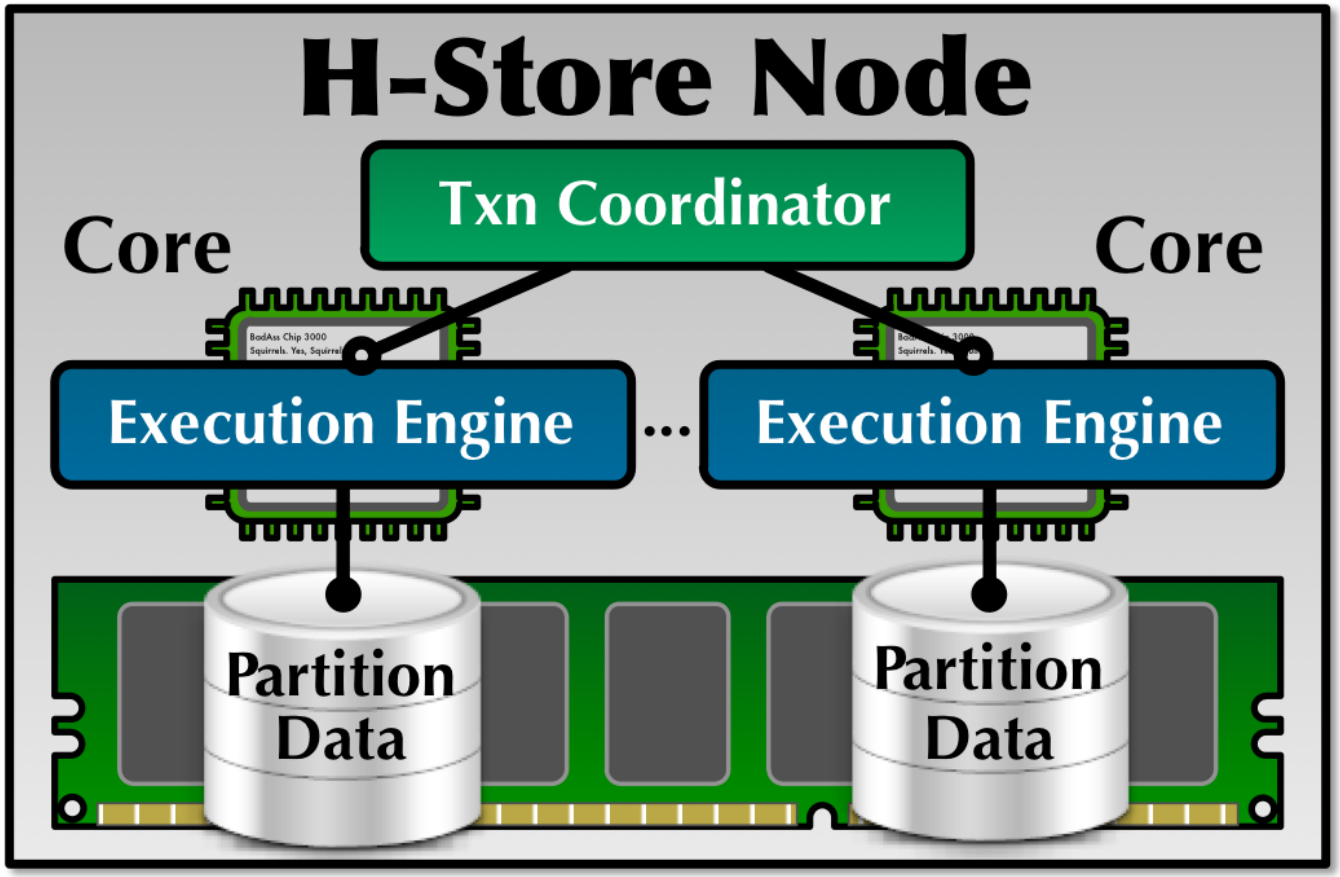
H-Store

Main Memory • Parallel • Shared-Nothing Transaction Processing

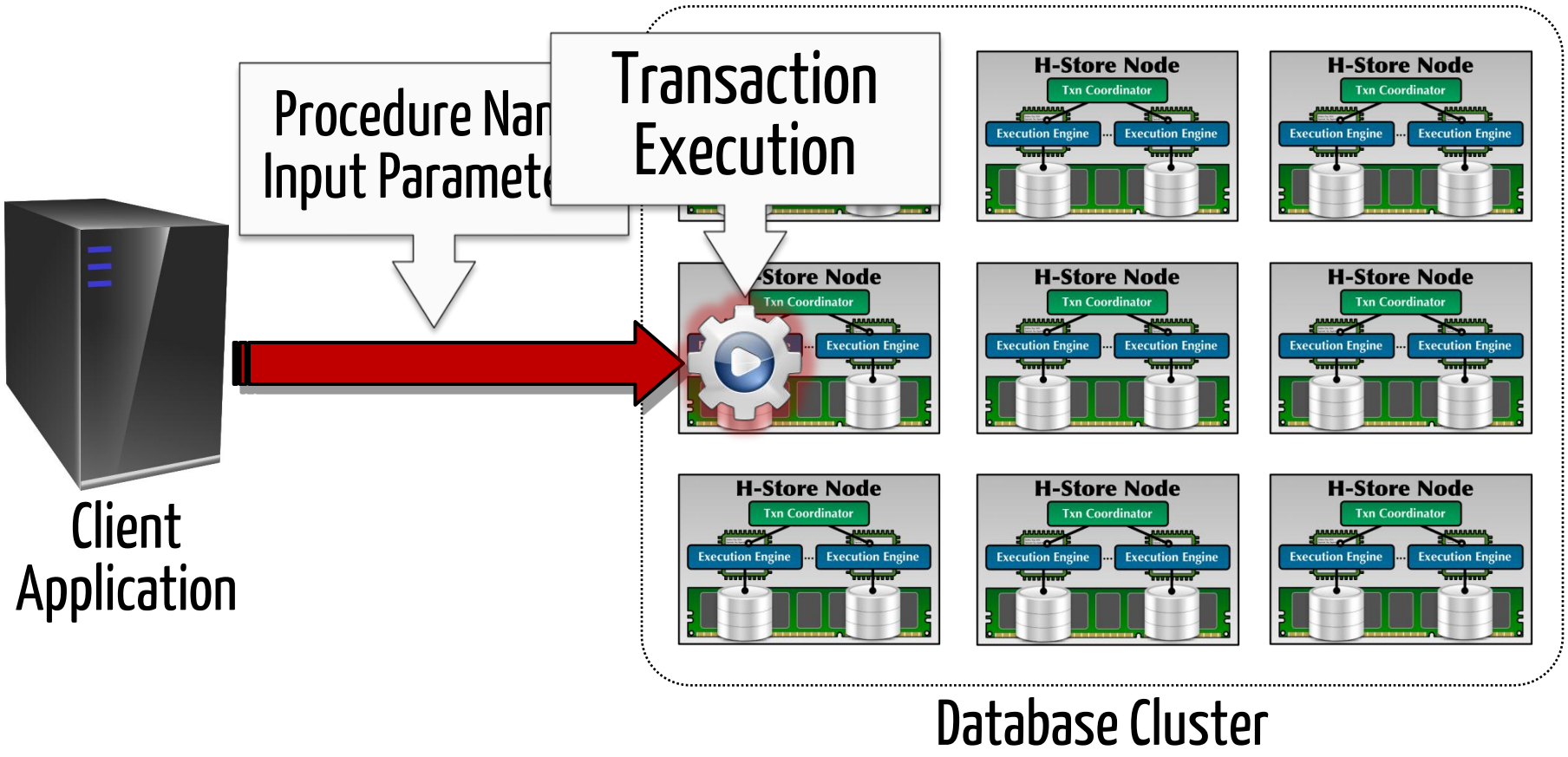
H-Store: A High-Performance, Distributed Main Memory Transaction Processing System

Proc. VLDB Endow., vol. 1, iss. 2, pp. 1496-1499, 2008.

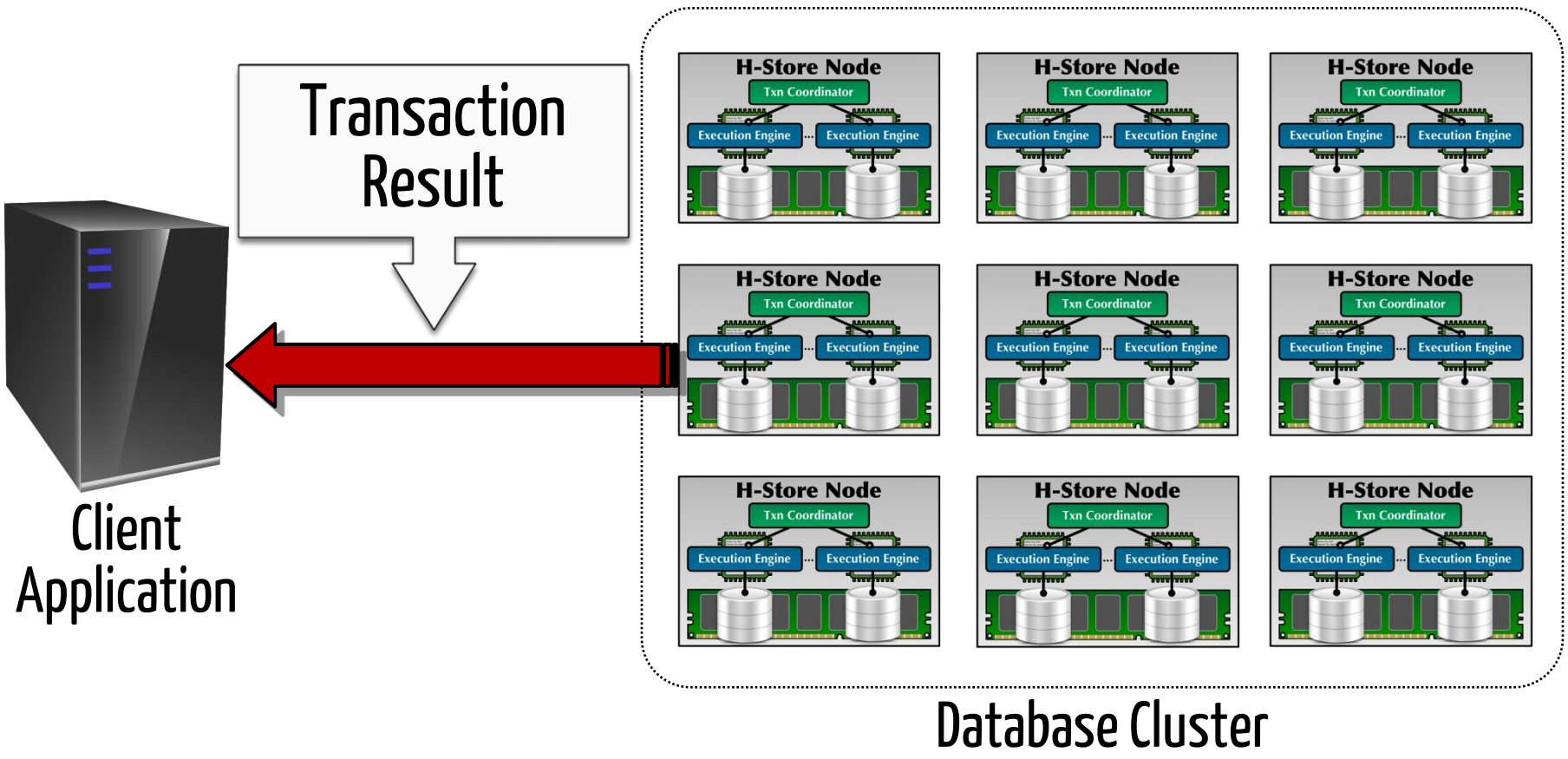
H-Store


H-Store Node
Txn Coordinator
Core
Core
Execution Engine
...
Execution Engine
Partition Data
Partition Data

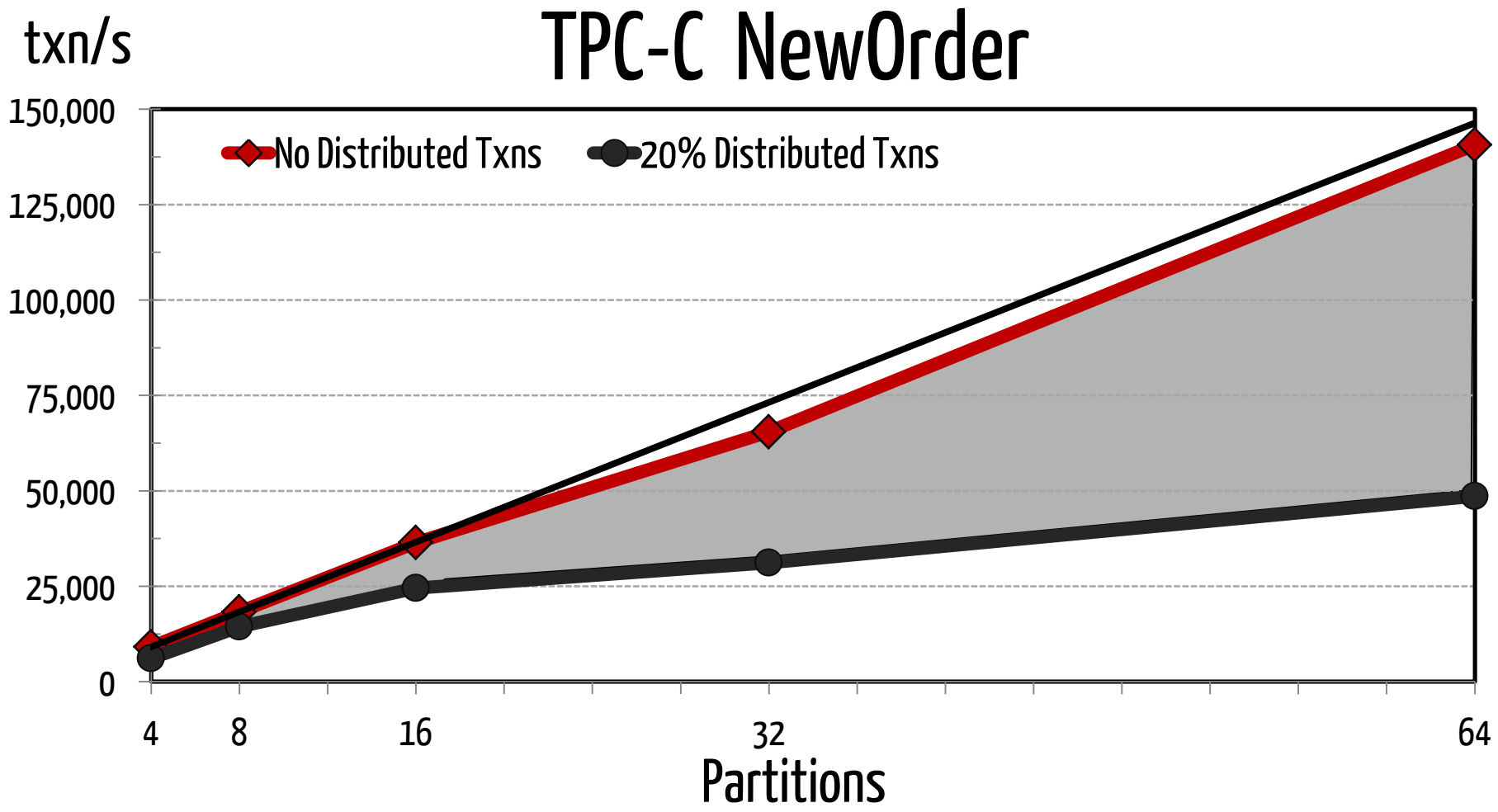
H-Store


Procedure Nar
Input Paramete
Transaction Execution
Client Application
H-Store Node
Txn Coordinator
Execution Engine
Database Cluster

H-Store


Transaction Result
Client Application
H-Store Node
Txn Coordinator
Execution Engine
Execution Engine
Database Cluster

H-Store


TPC-C NewOrder
txn/s
150,000
125,000
100,000
75,000
50,000
25,000
0
No Distributed Txns
20% Distributed Txns
4
8
16
32
64
Partitions

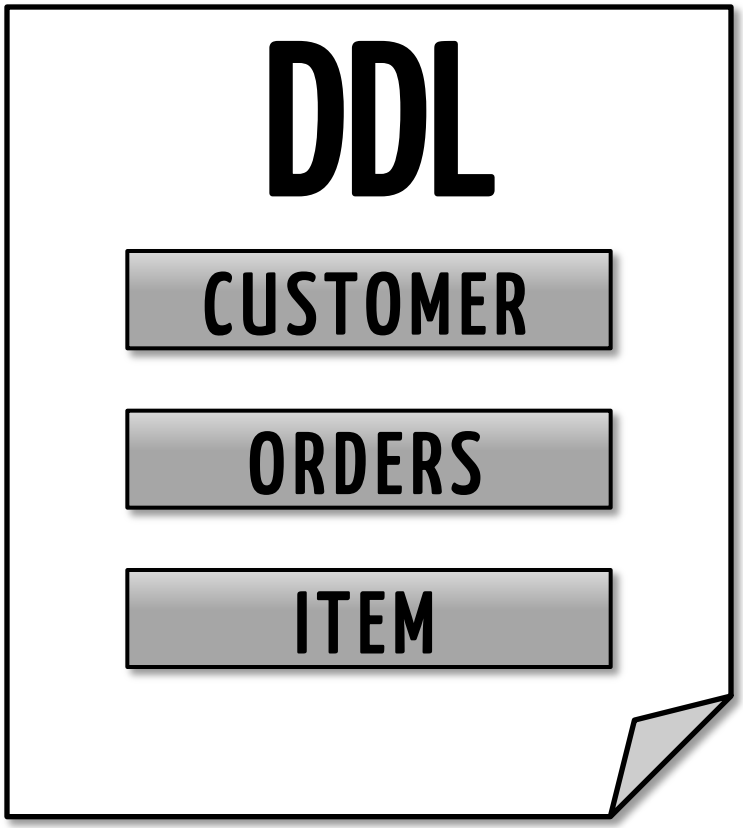
Horticulture

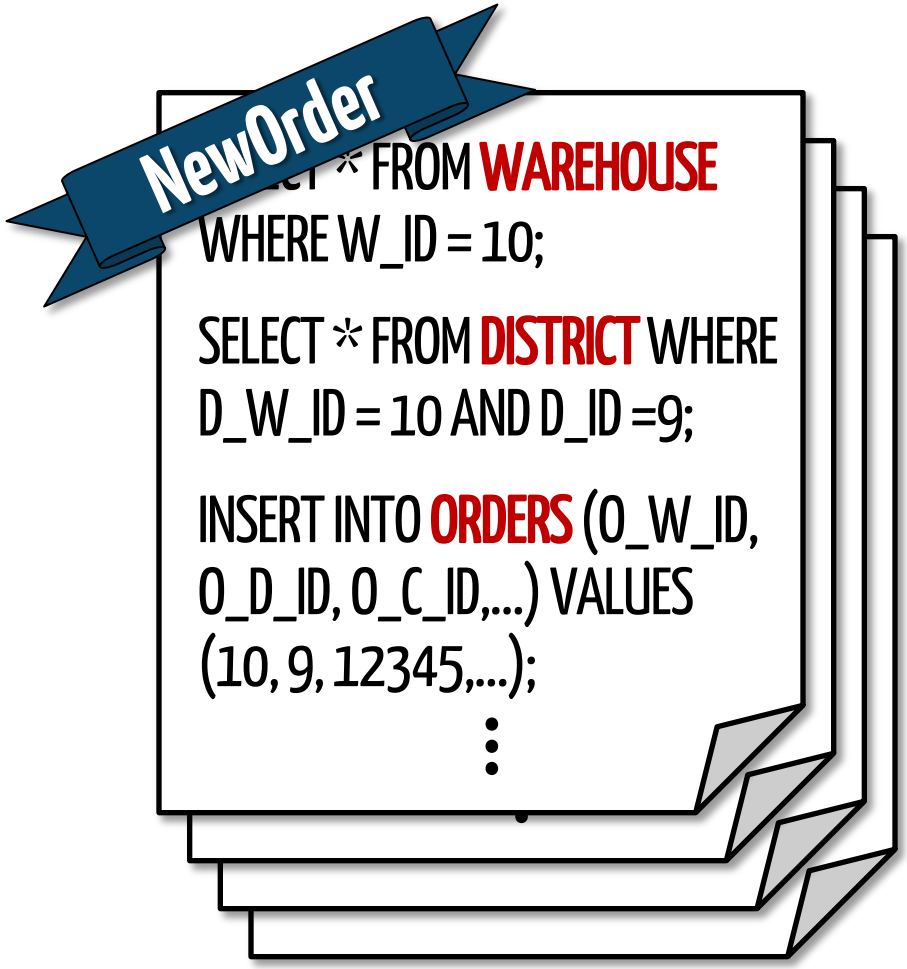
Automatic Database Design Tool for Parallel Systems

Skew-Aware Automatic Database Partitioning in Shared-Nothing, Parallel OLTP Systems

SIGMOD 2012


Horticulture


DDL
CUSTOMER
ORDERS
ITEM


NewOrder
* FROM WAREHOUSE
WHERE W_ID = 10;
SELECT * FROM DISTRICT WHERE
D_W_ID = 10 AND D_ID =9;
INSERT INTO ORDERS (O_W_ID,
O_D_ID, O_C_ID,...) VALUES
(10, 9, 12345,...);

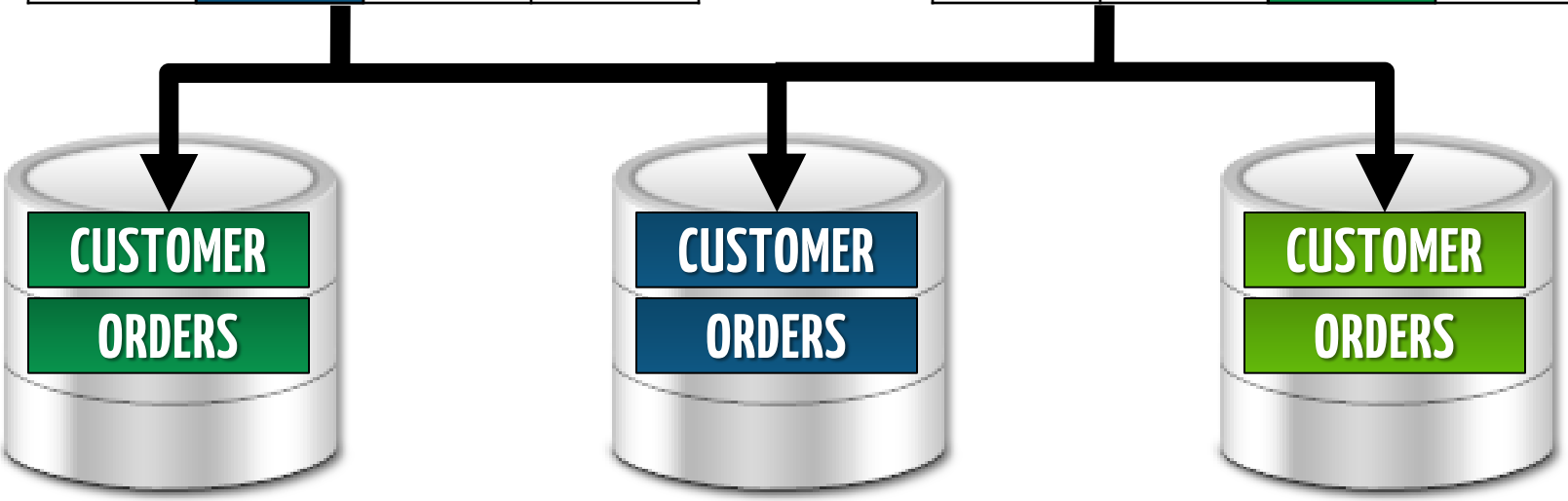
Horticulture

CUSTOMER

c_id	c_w_id	c_last	...
1001	5	RZA	-
1002	3	GZA	-
1003	12	Raekwon	-
1004	5	Deck	-
1005	6	Killah	-
1006	7	ODB	-

ORDERS

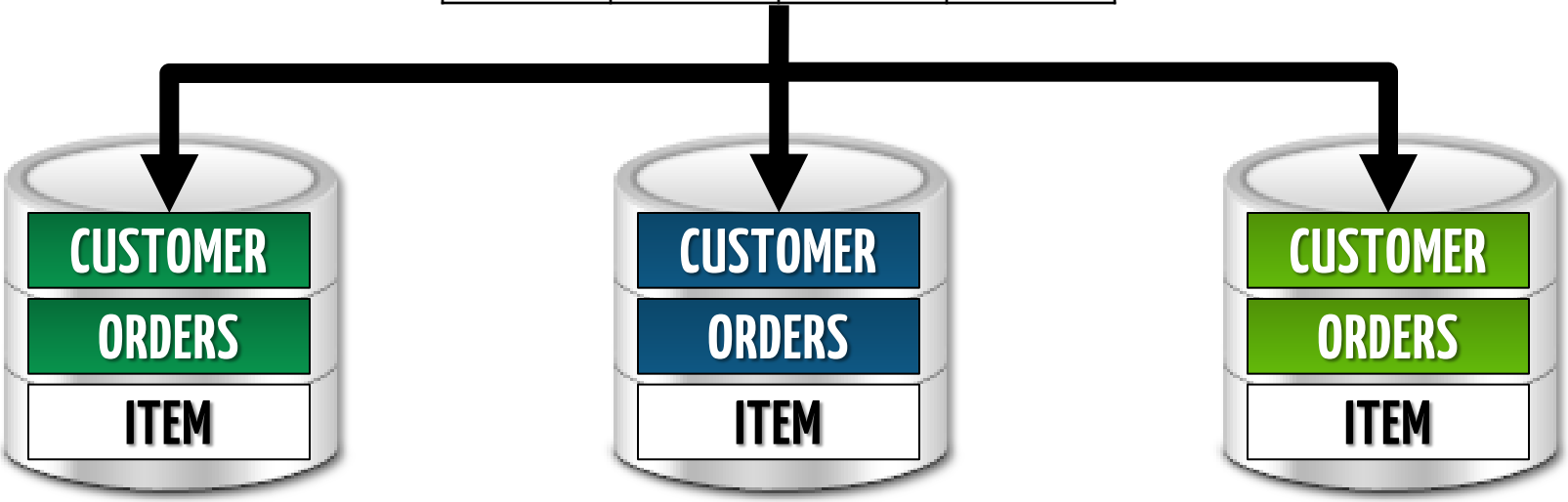
o_id	o_c_id	o_w_id	...
78703	1004	5	-
78704	1002	3	-
78705	1006	7	-
78706	1005	6	-
78707	1005	6	-
78708	1003	12	-



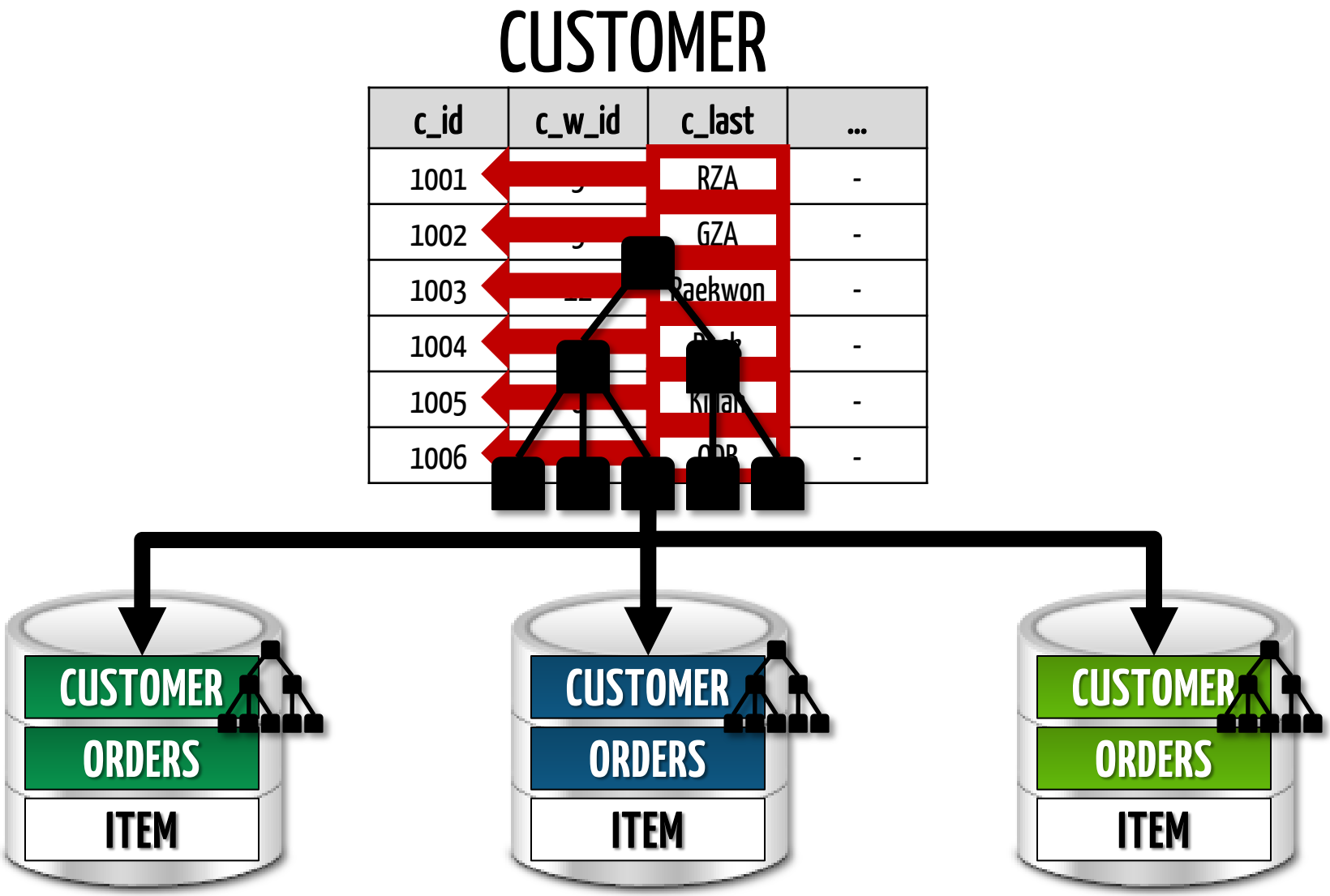
Horticulture

ITEM

i_id	i_name	i_price	...
603514	XXX	23.99	-
267923	XXX	19.99	-
475386	XXX	14.99	-
578945	XXX	9.98	-
476348	XXX	103.49	-
784285	XXX	69.99	-



Horticulture



Horticulture

Client Application

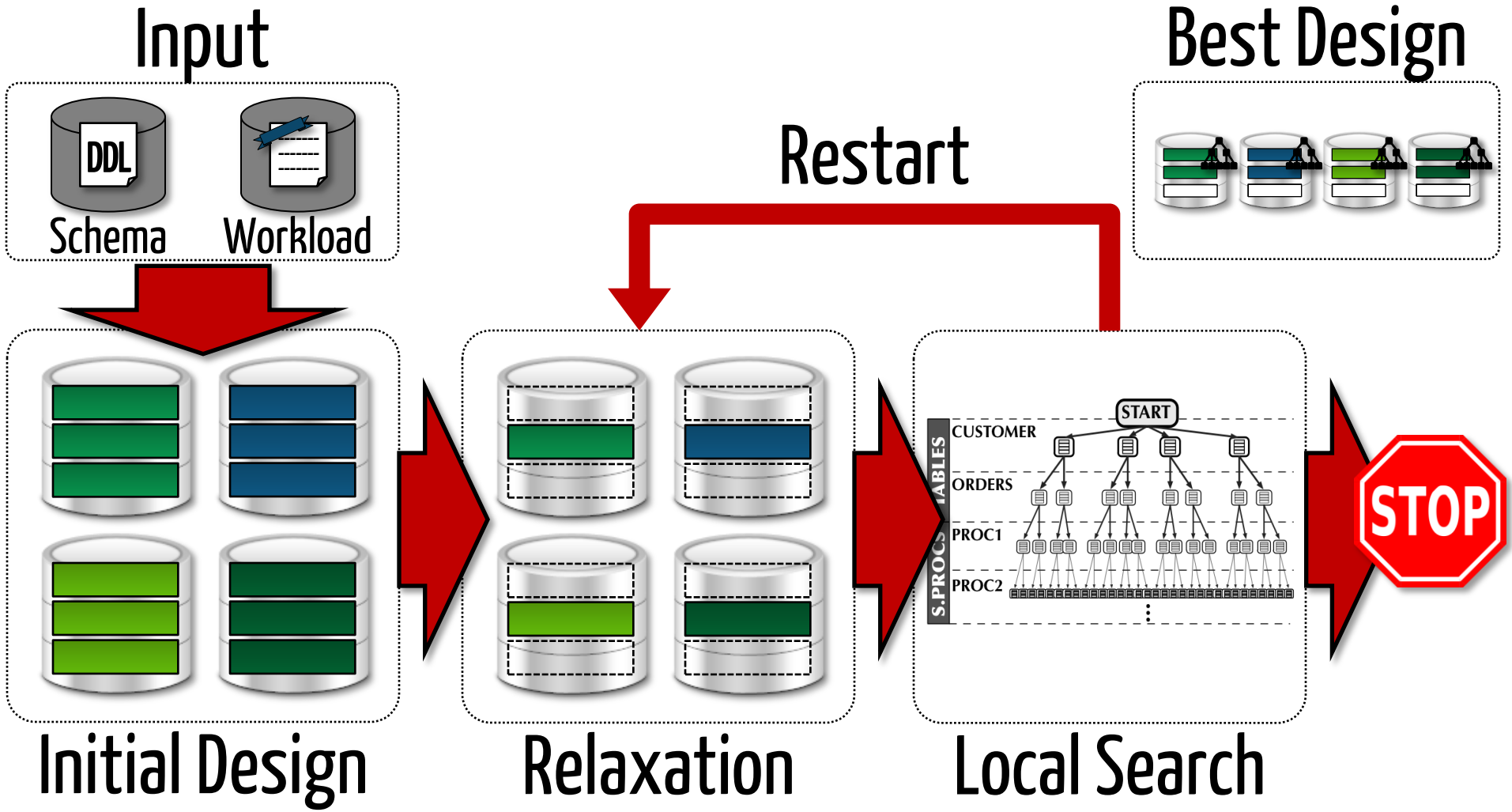
NewOrder(5, “Method Man”, 1234)

Horticulture

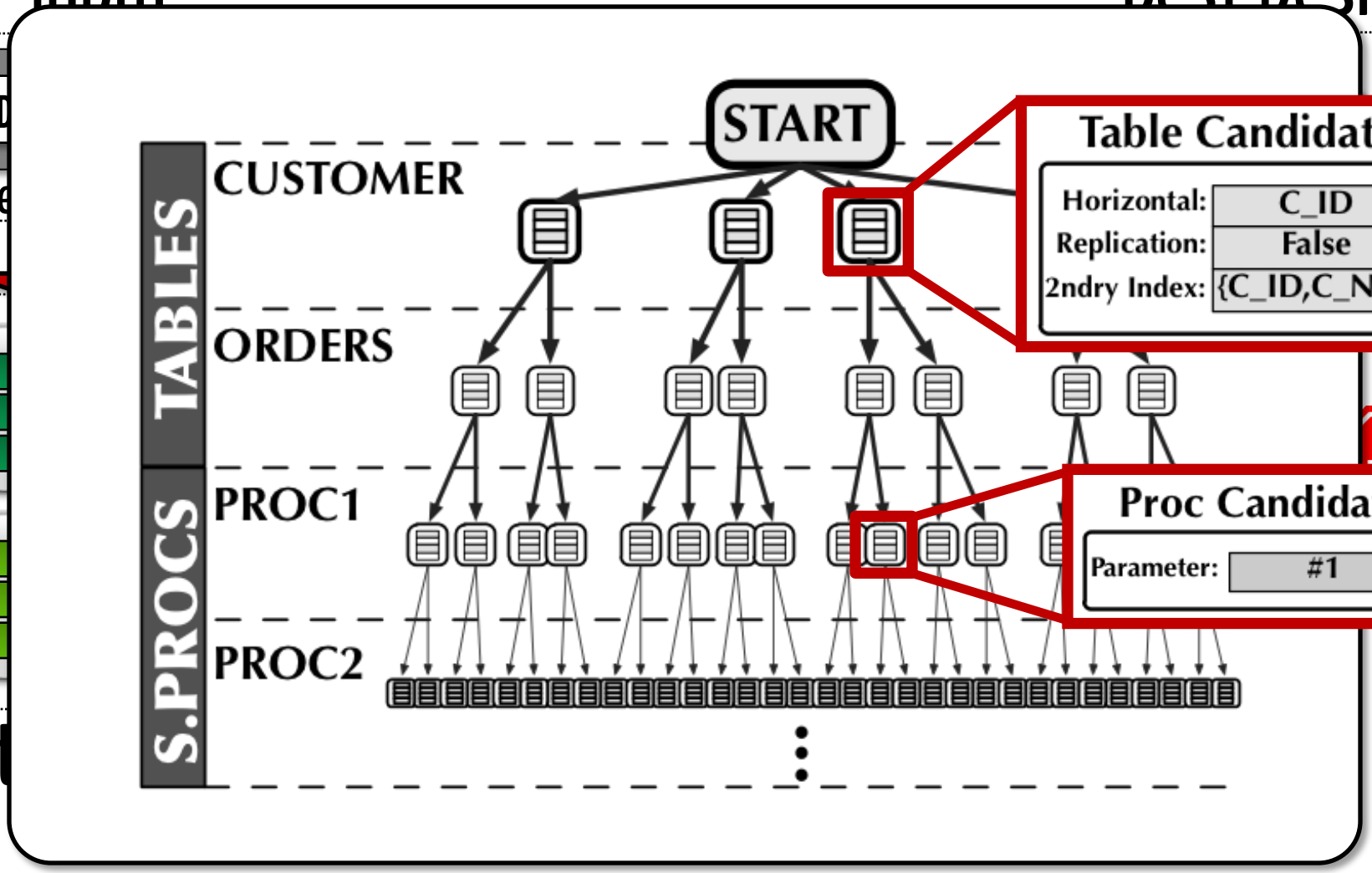
Client Application

NewOrder(5, "Method Man", 1234)

Large-Neighborhood Search

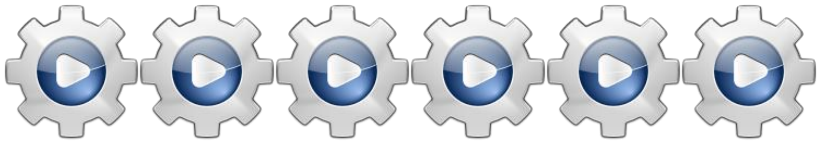


Large-Neighborhood Search

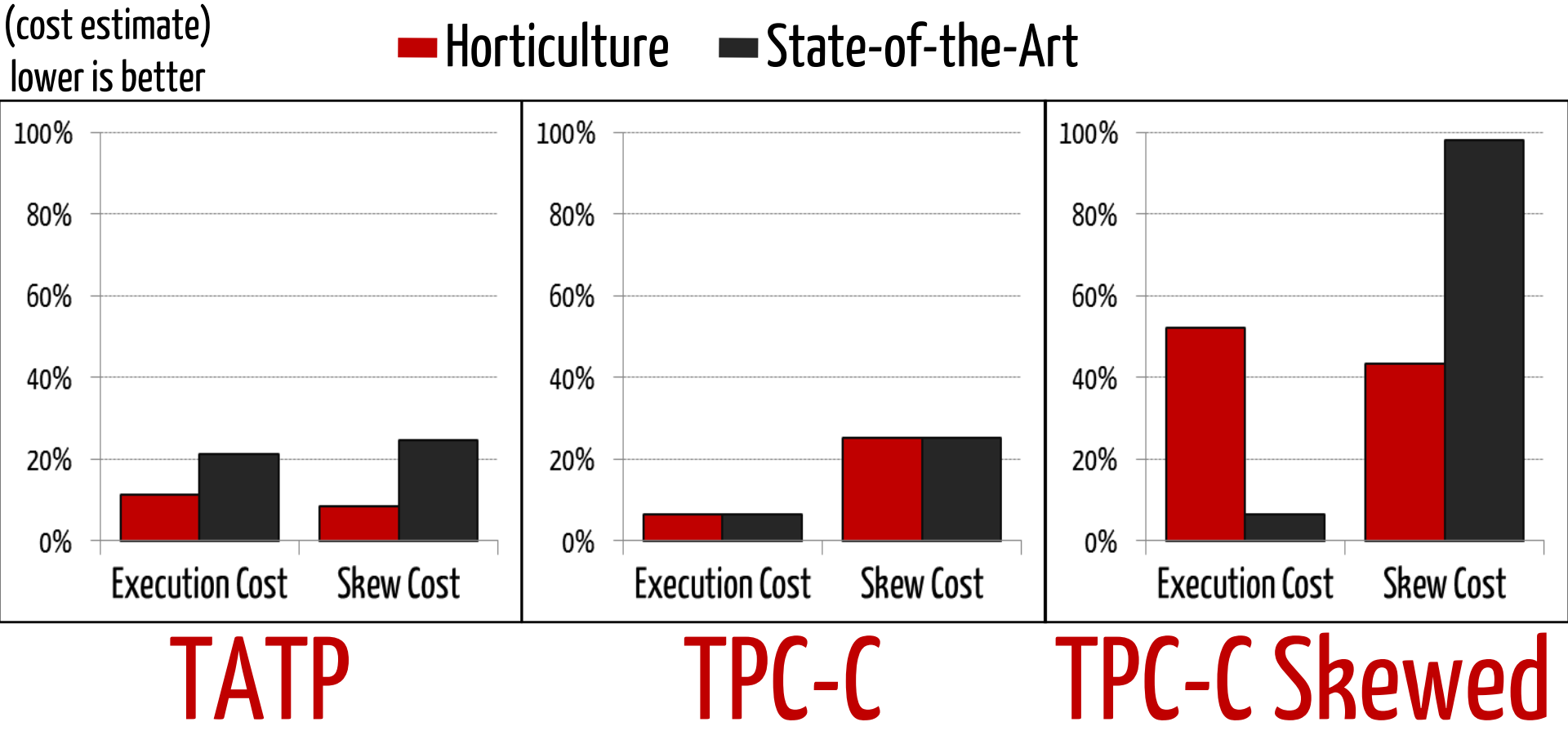


Cost Model

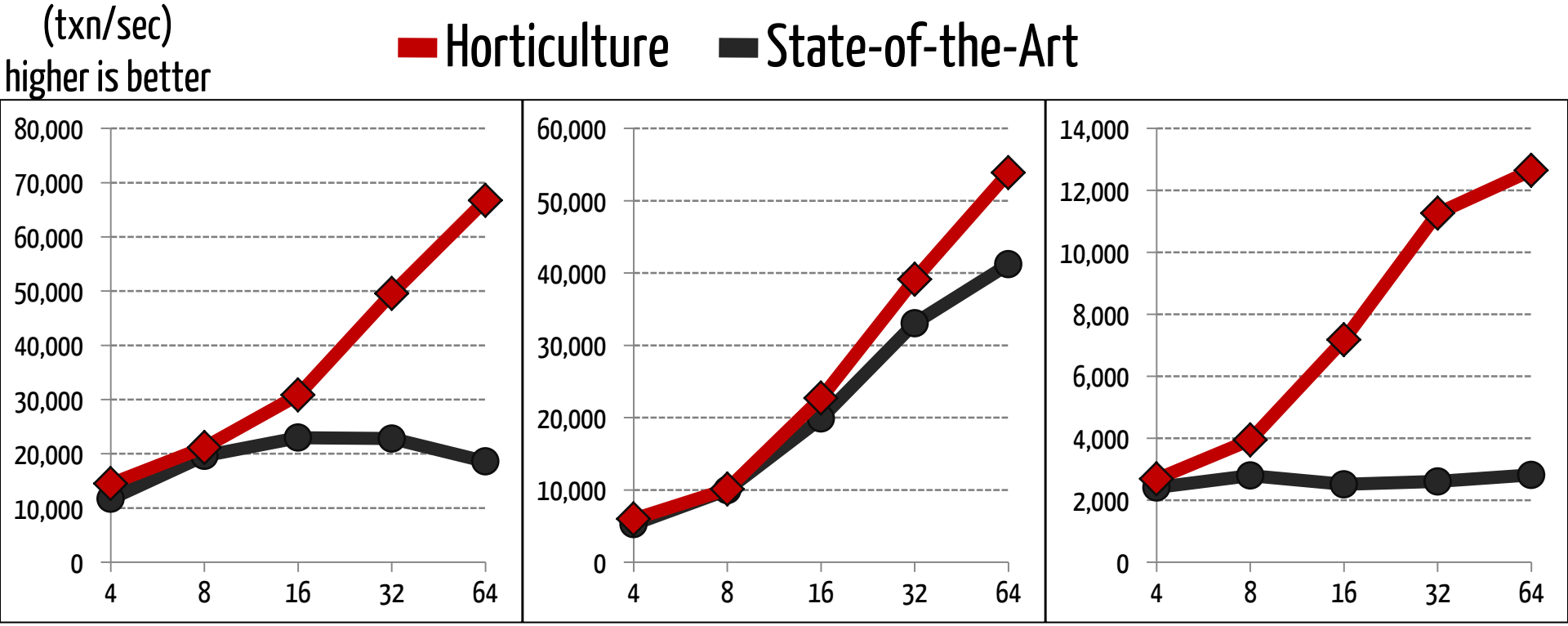
Distributed Transactions + Workload Skew Factor



Algorithm Comparison



Throughput



TATP
+88%

TPC-C
+16%

TPC-C Skewed
+183%

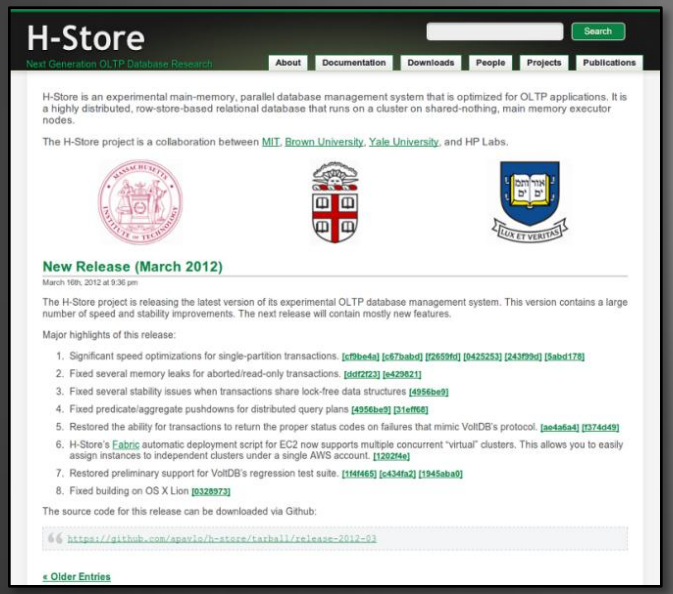

-Store

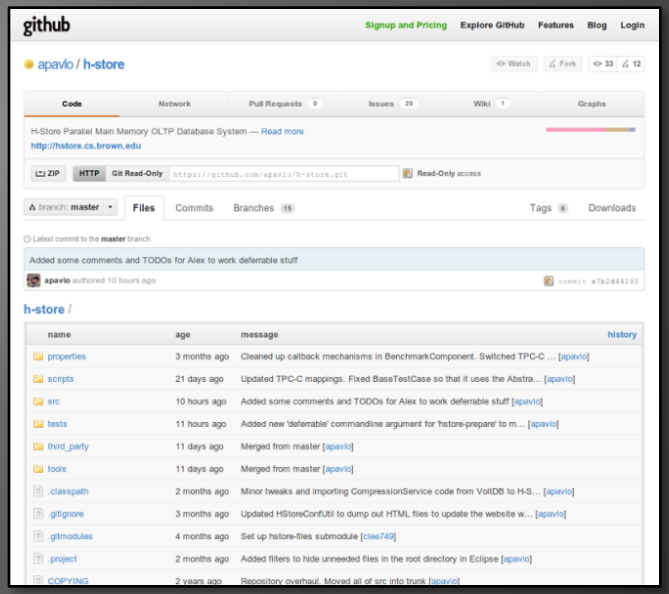
Conclusion:

Dating scene is still difficult.

But partitioning your database is now easier.

H-Store





http://hstore.cs.brown.edu

http://github.com/apavlo/h-store