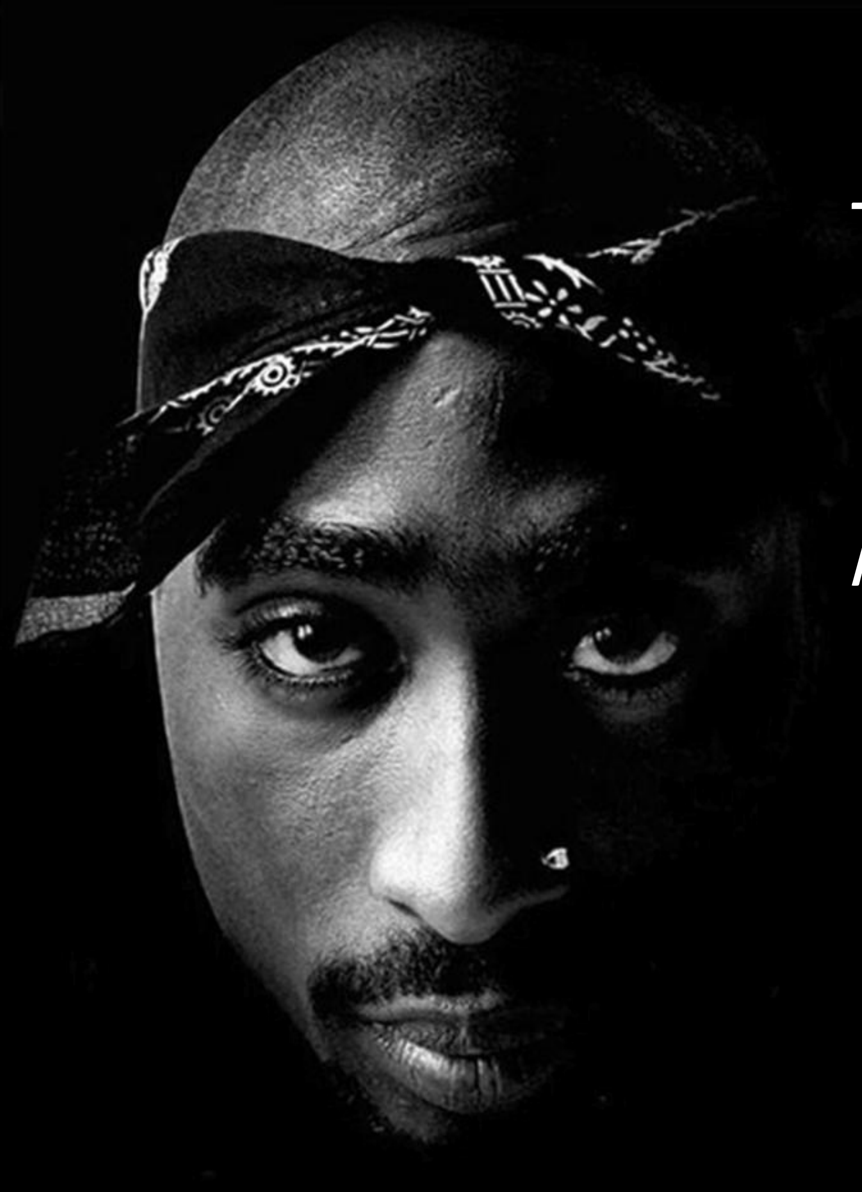
Methods for the Optimization of Parallel OLTP Systems

@andy_pavlo

BROWN

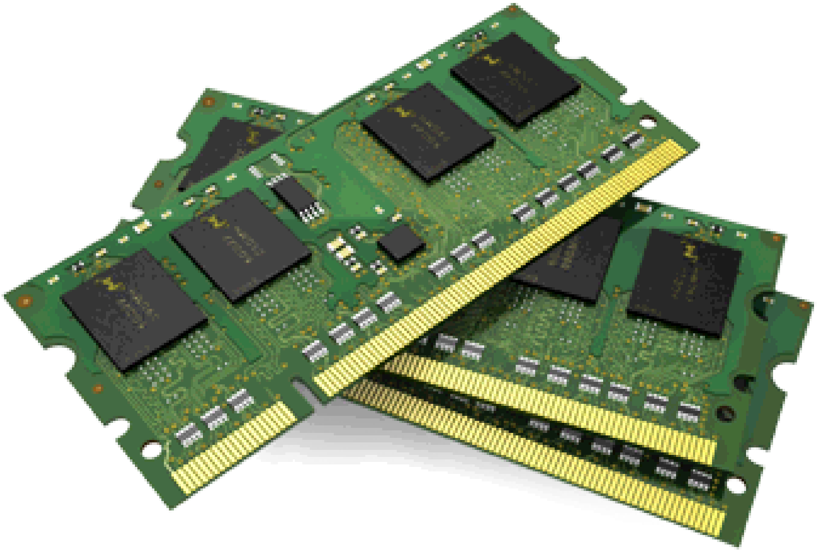
Thesis Proposal
May 2012



Tellin' him just be ready set
Pack ya shit up quick;

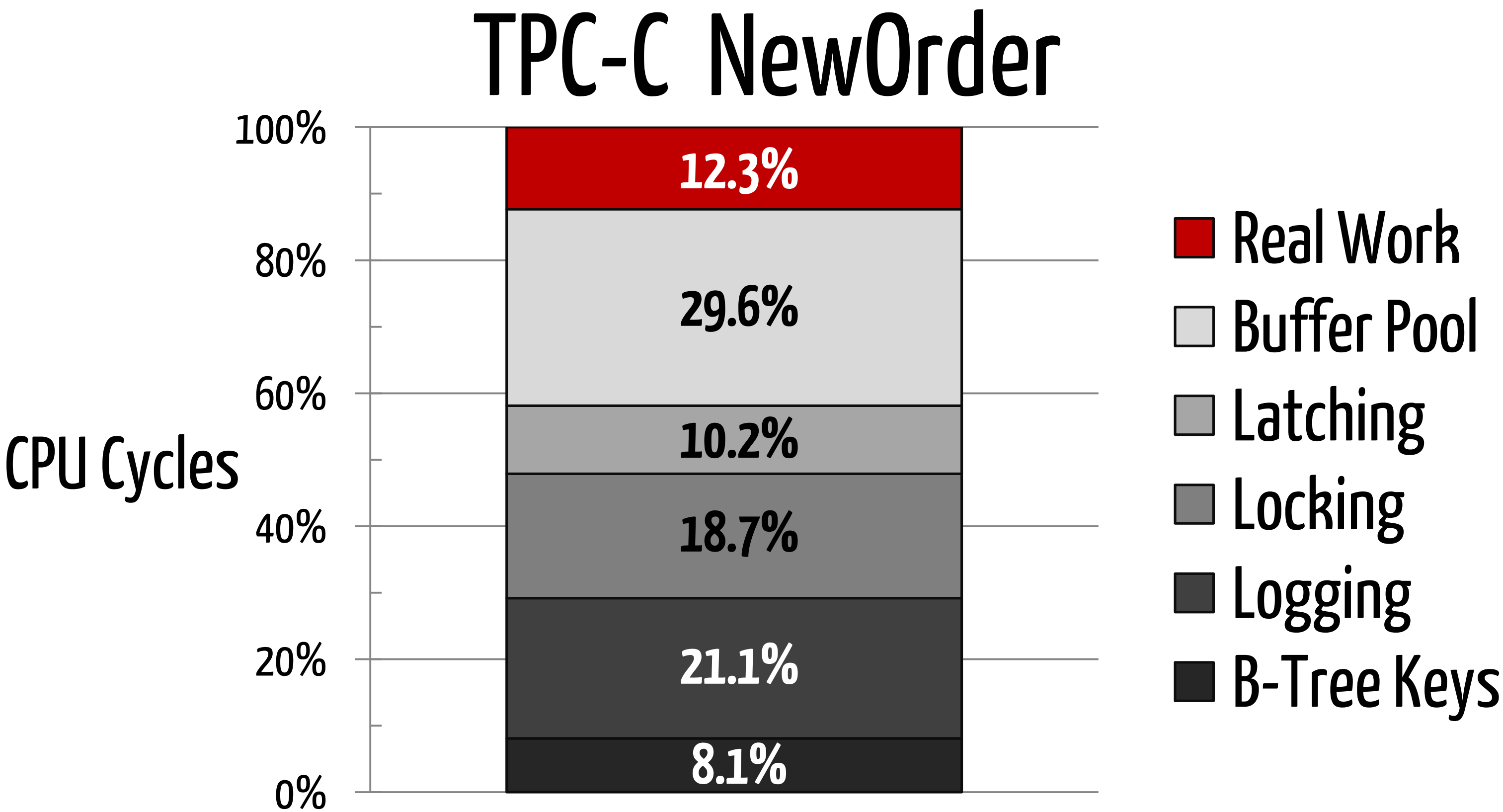
And when I hit,
be prepared to jet

Soulja's Story (1991)



Fast
+
Cheap

Legacy Systems



OLTP Through the Looking Glass, and What We Found There
SIGMOD 2008

OLTP Transactions

Fast Repetitive Small

H-Store

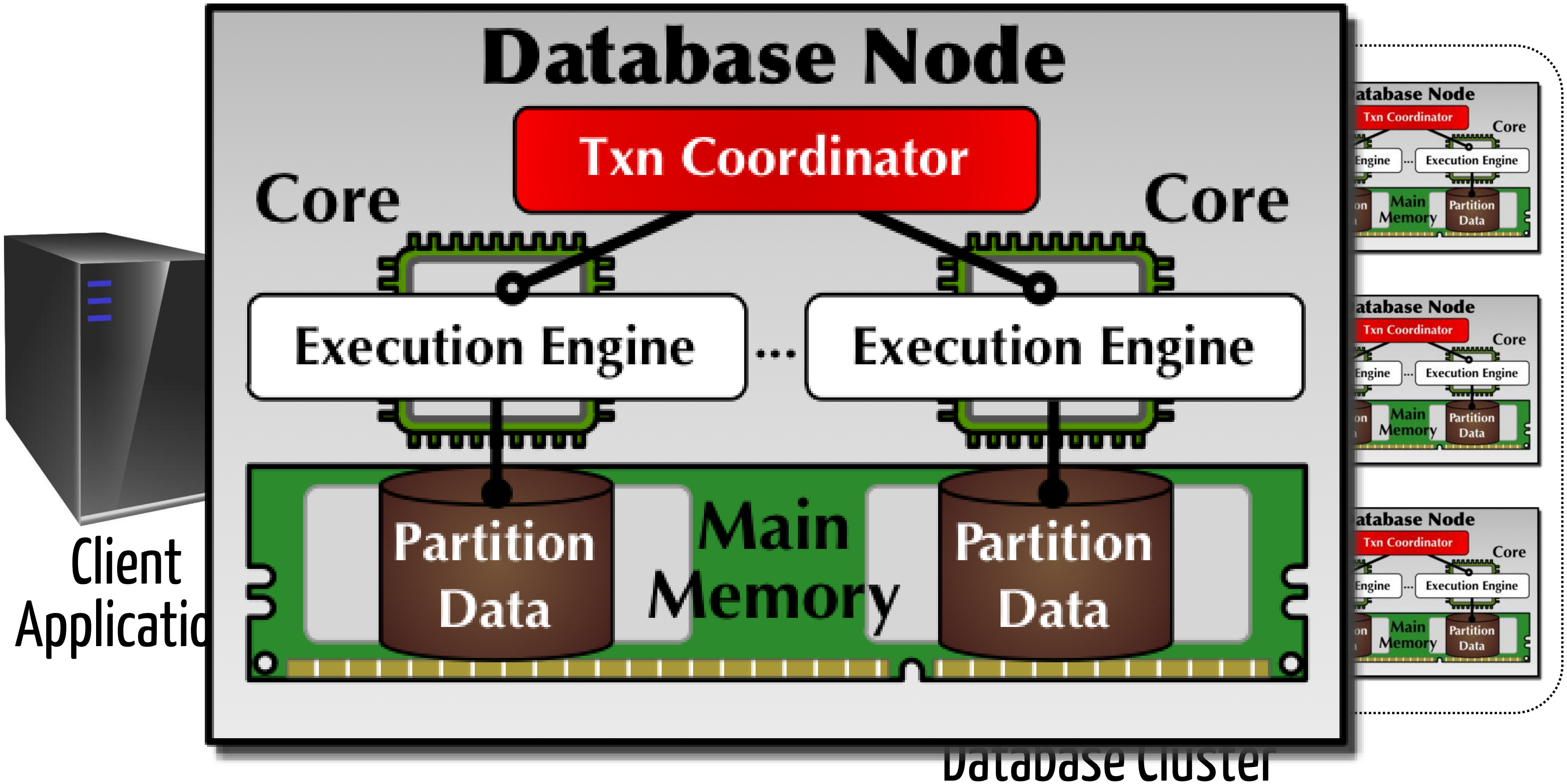
Main Memory • Parallel • Shared-Nothing Transaction Processing

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

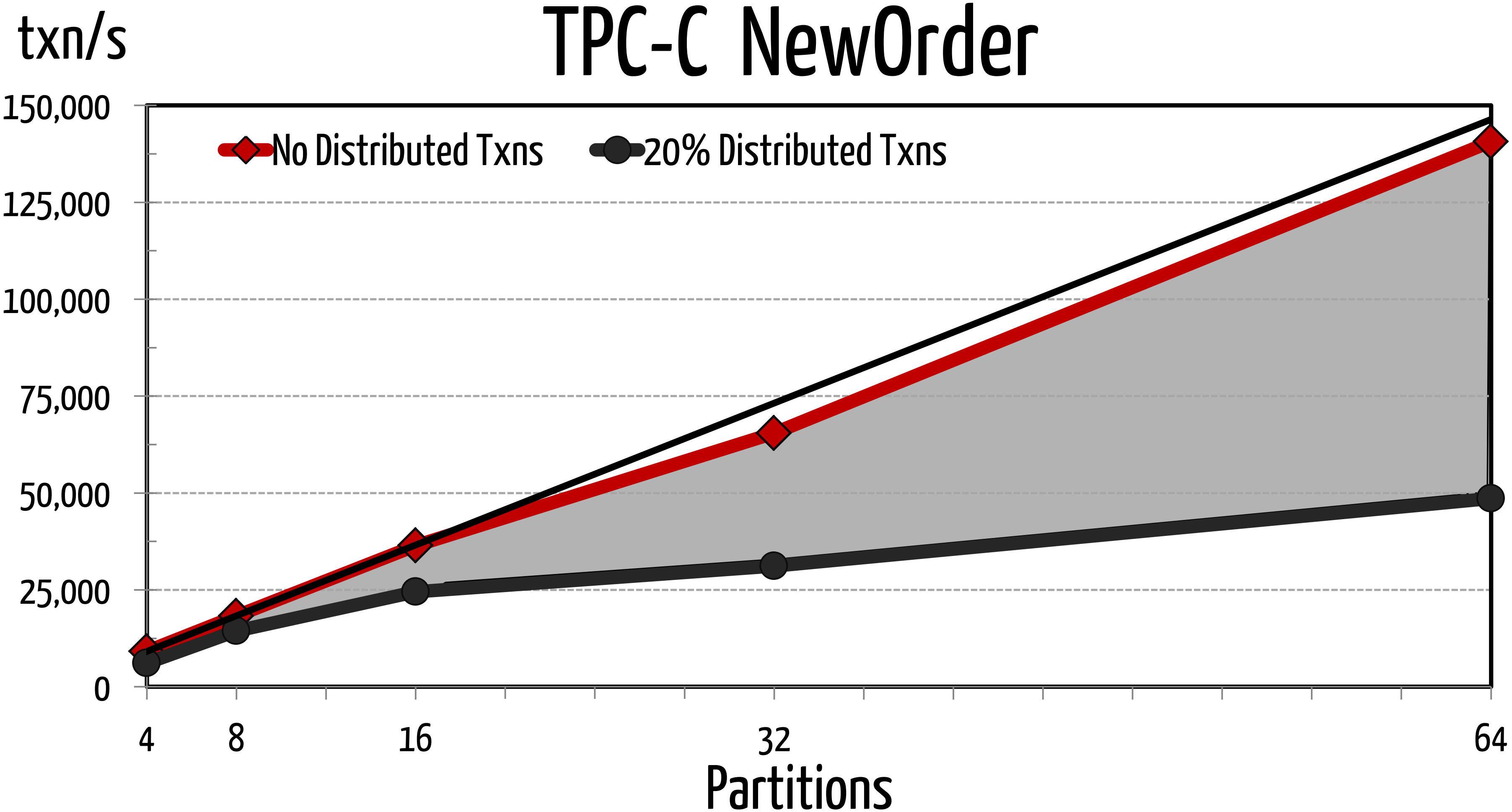
Proc. VLDB Endow., vol. 1, iss. 2, pp. 1496-1499, 2008.

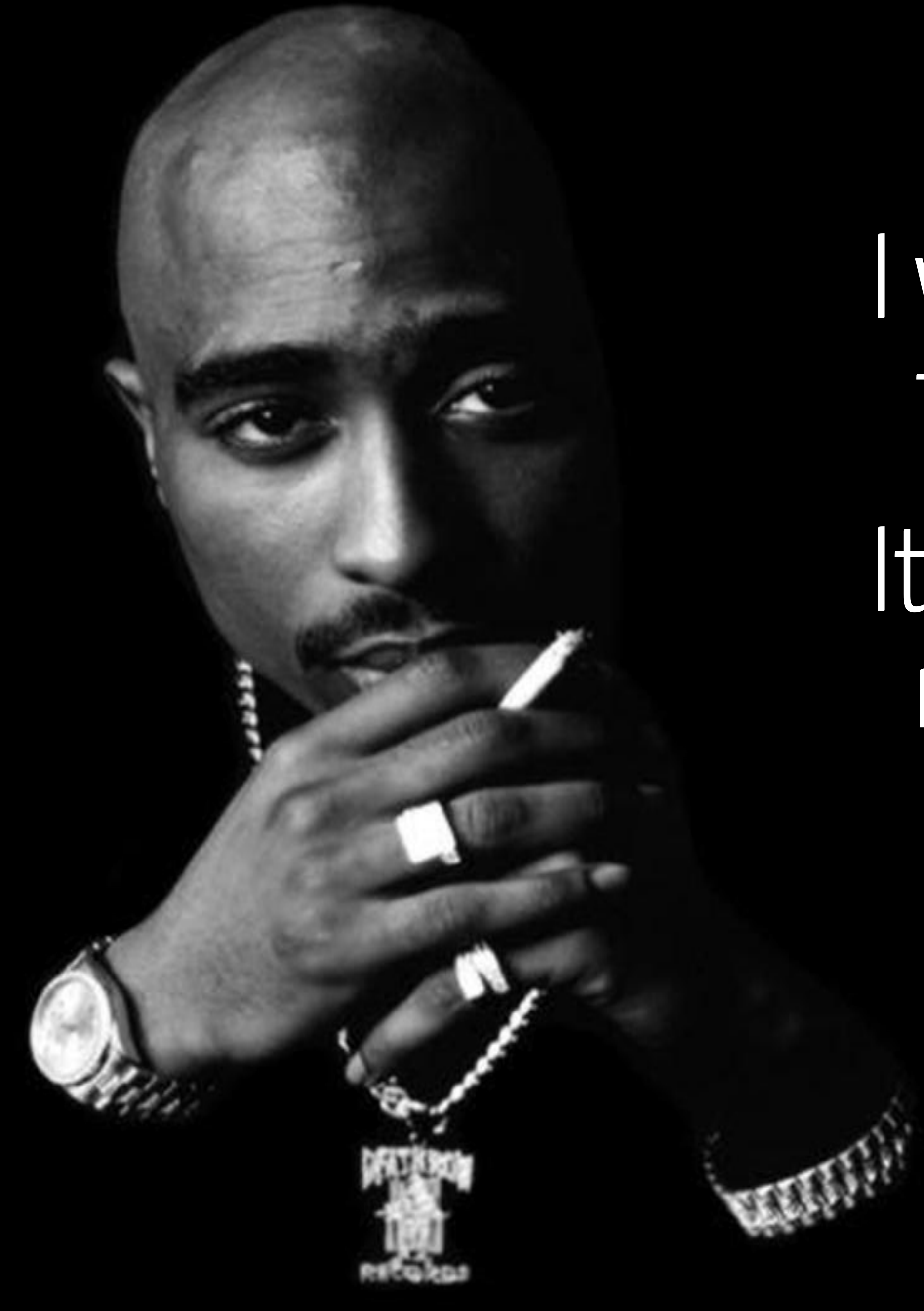

H-Store: A High-Performance, Distributed Main Memory Transaction Processing System

H-Store


Database Node
Txn Coordinator
Core
Core
Execution Engine
Execution Engine
Partition Data
Main Memory
Partition Data
Client
Database Node
Txn Coordinator
Core
Execution Engine
Main Memory
Partition Data

TPC-C NewOrder





I wasn't mad until
these tricks played me,

It's time to sanitize
my posse.

Street Fame (1994)

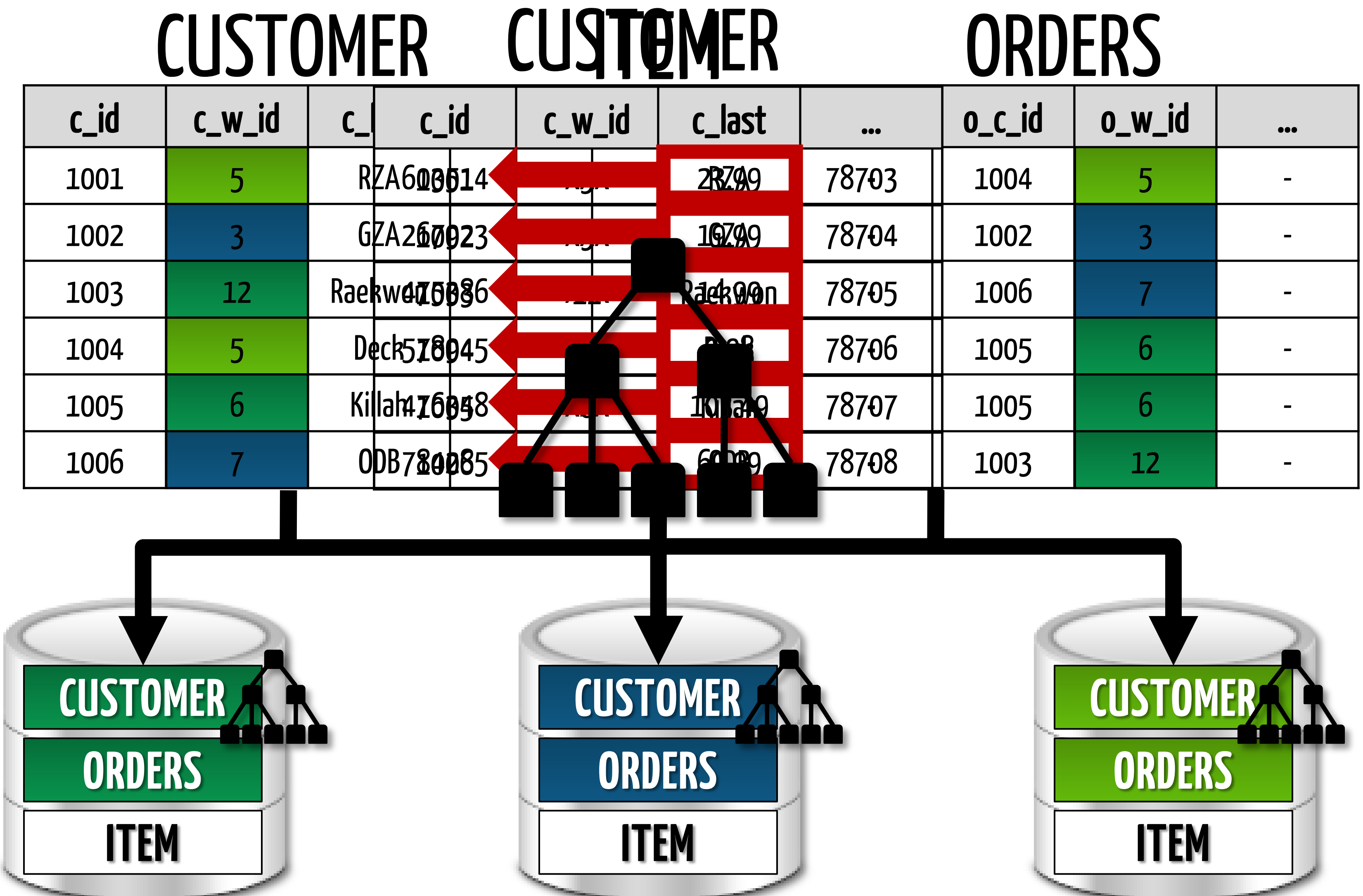
Optimization #1:

Partition database to reduce the number of distributed txns.

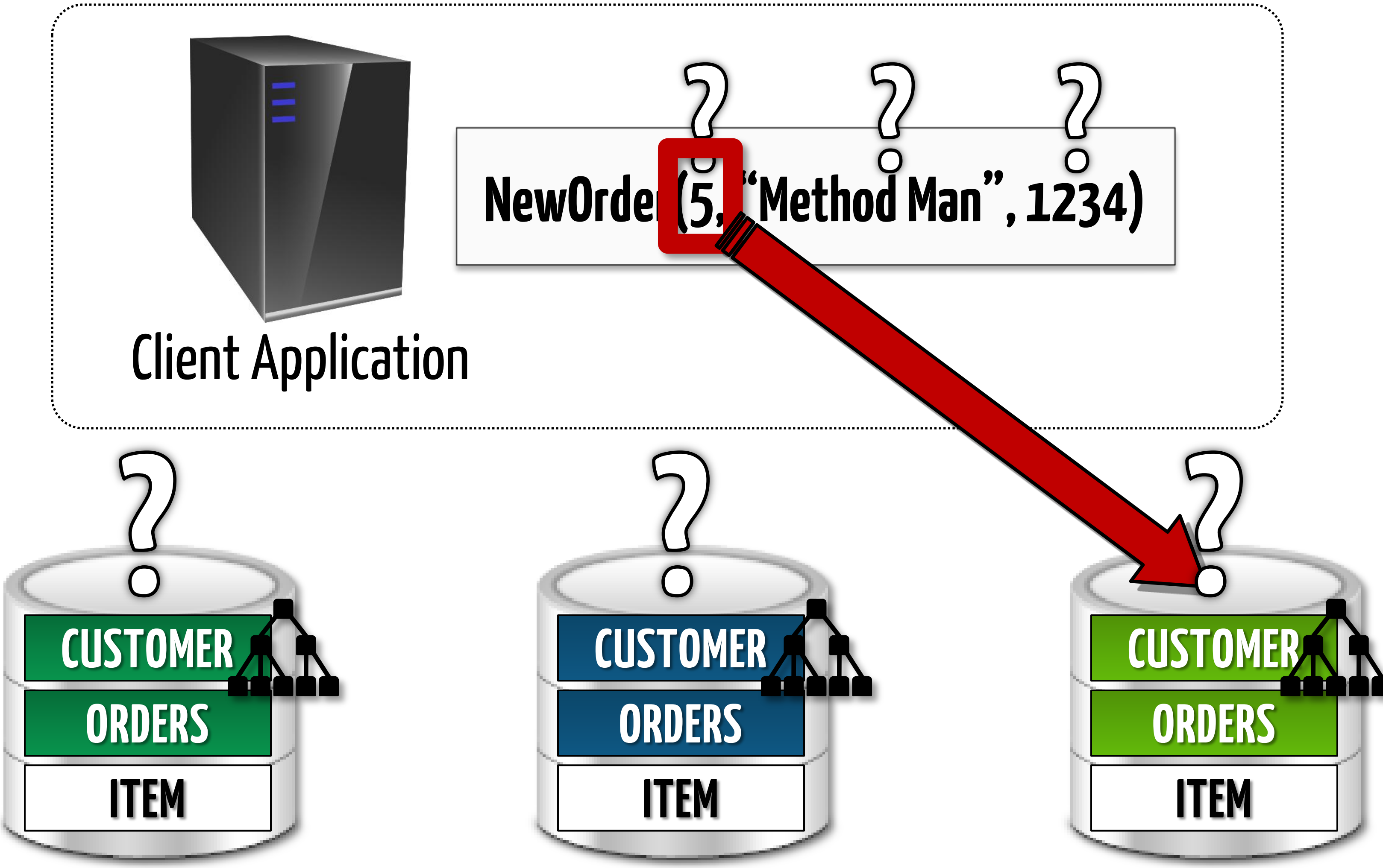
Skew-Aware Automatic Database Partitioning in Shared-Nothing, Parallel OLTP Systems
SIGMOD 2012


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

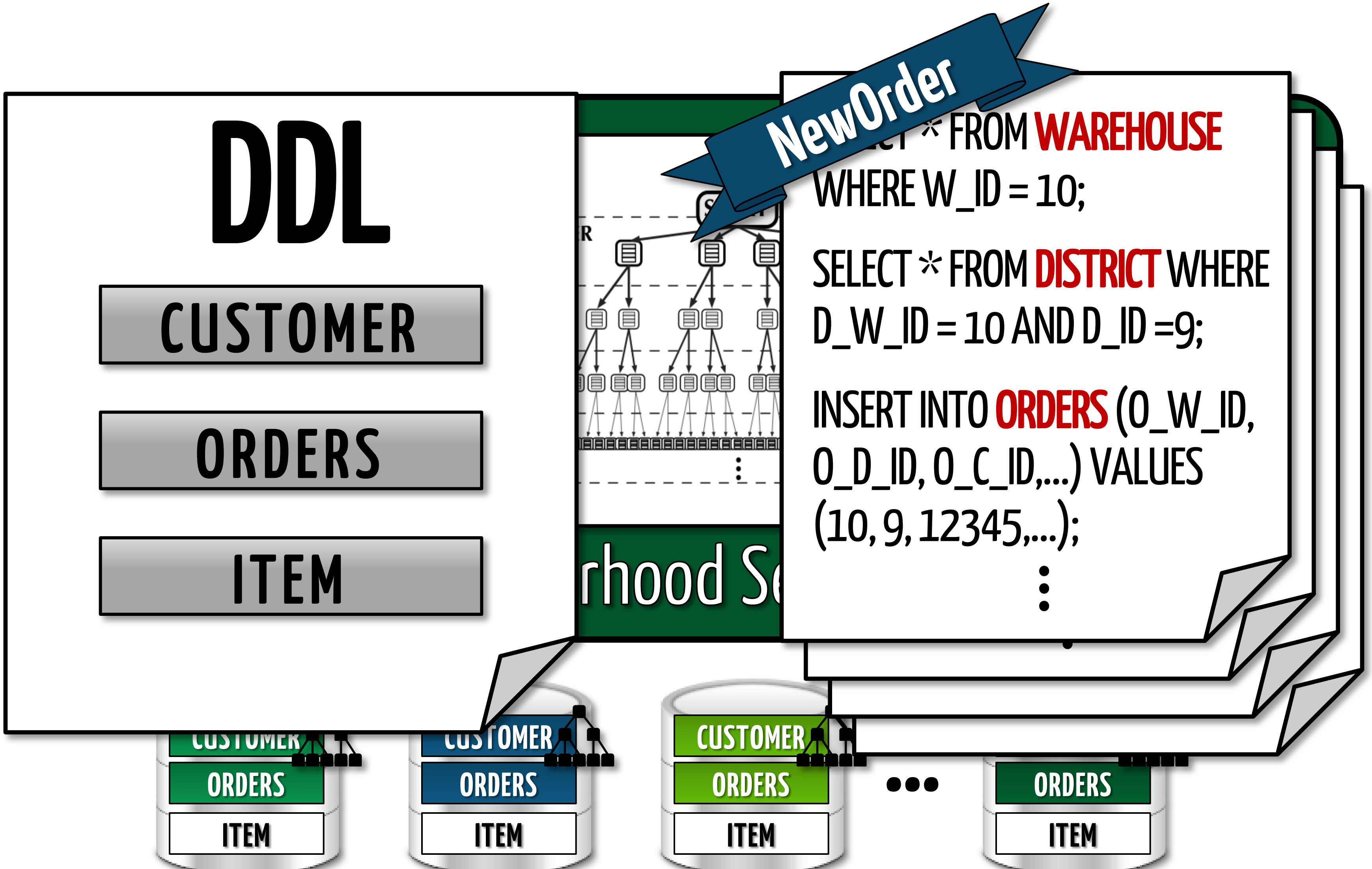
Horticulture



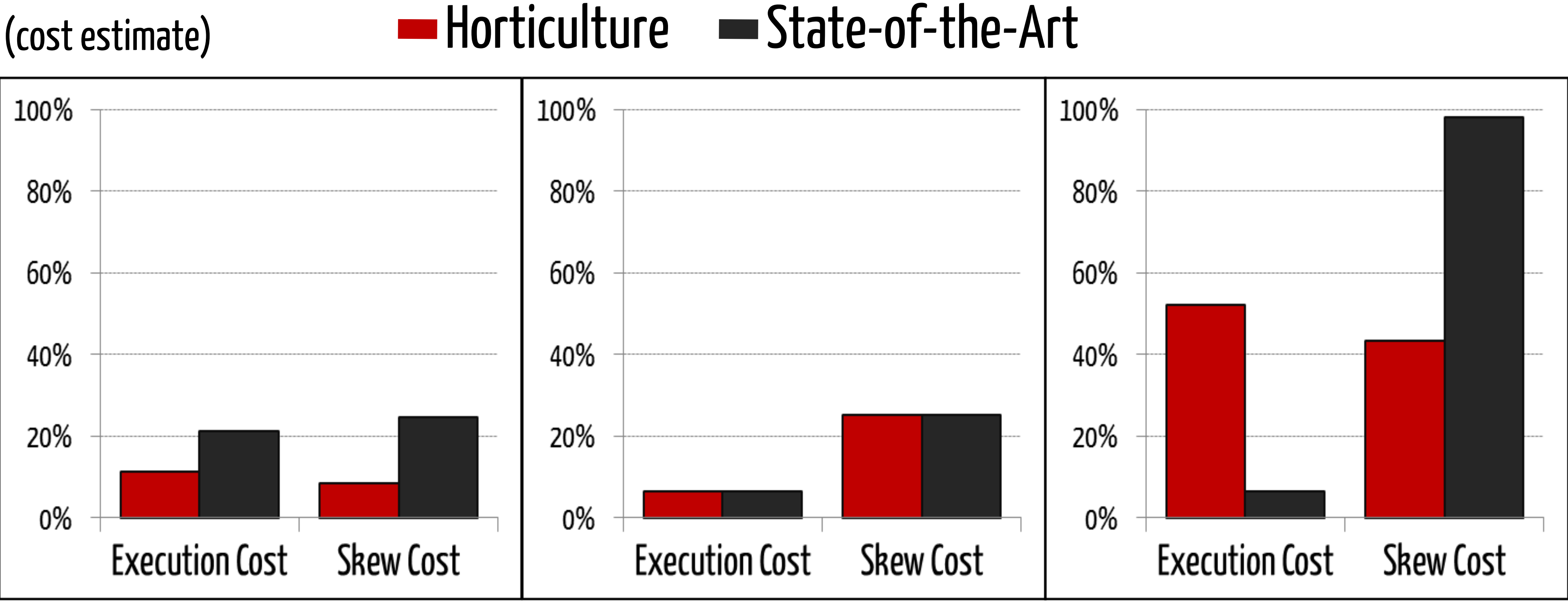
Horticulture



Horticulture

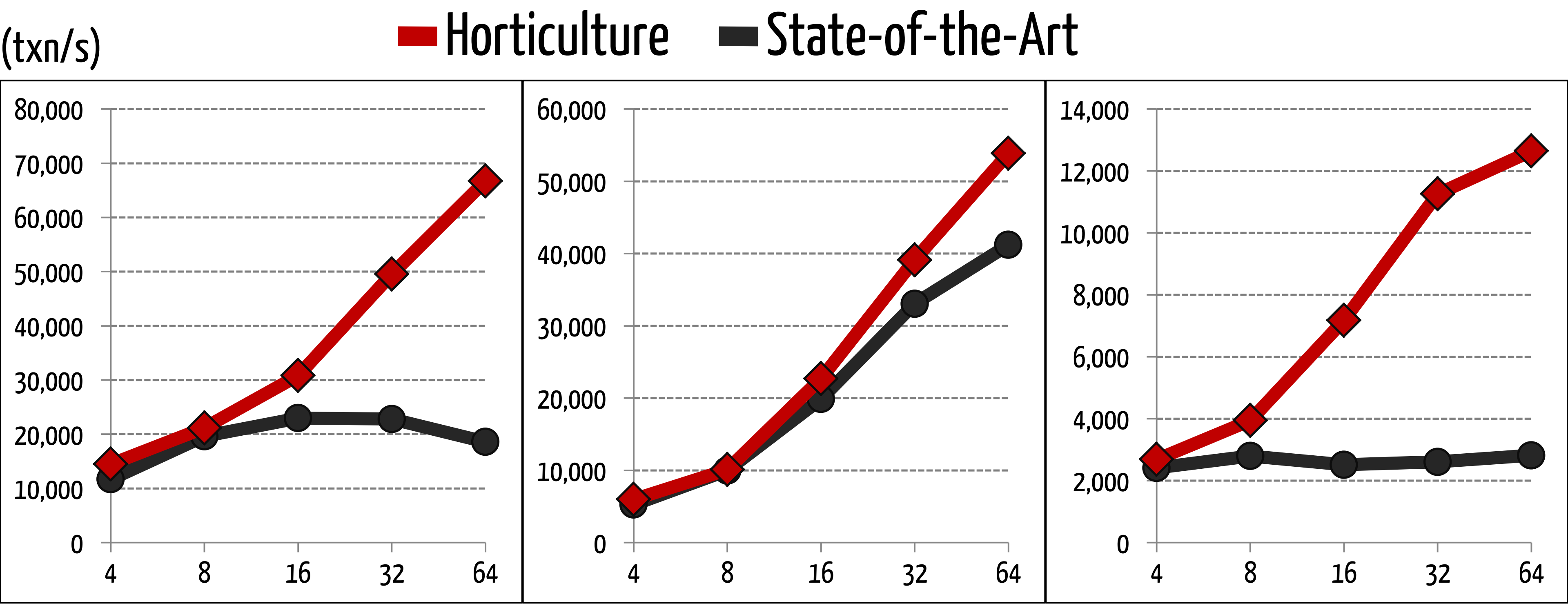


Algorithm Comparison



TATP TPC-C TPC-C Skewed

Throughput

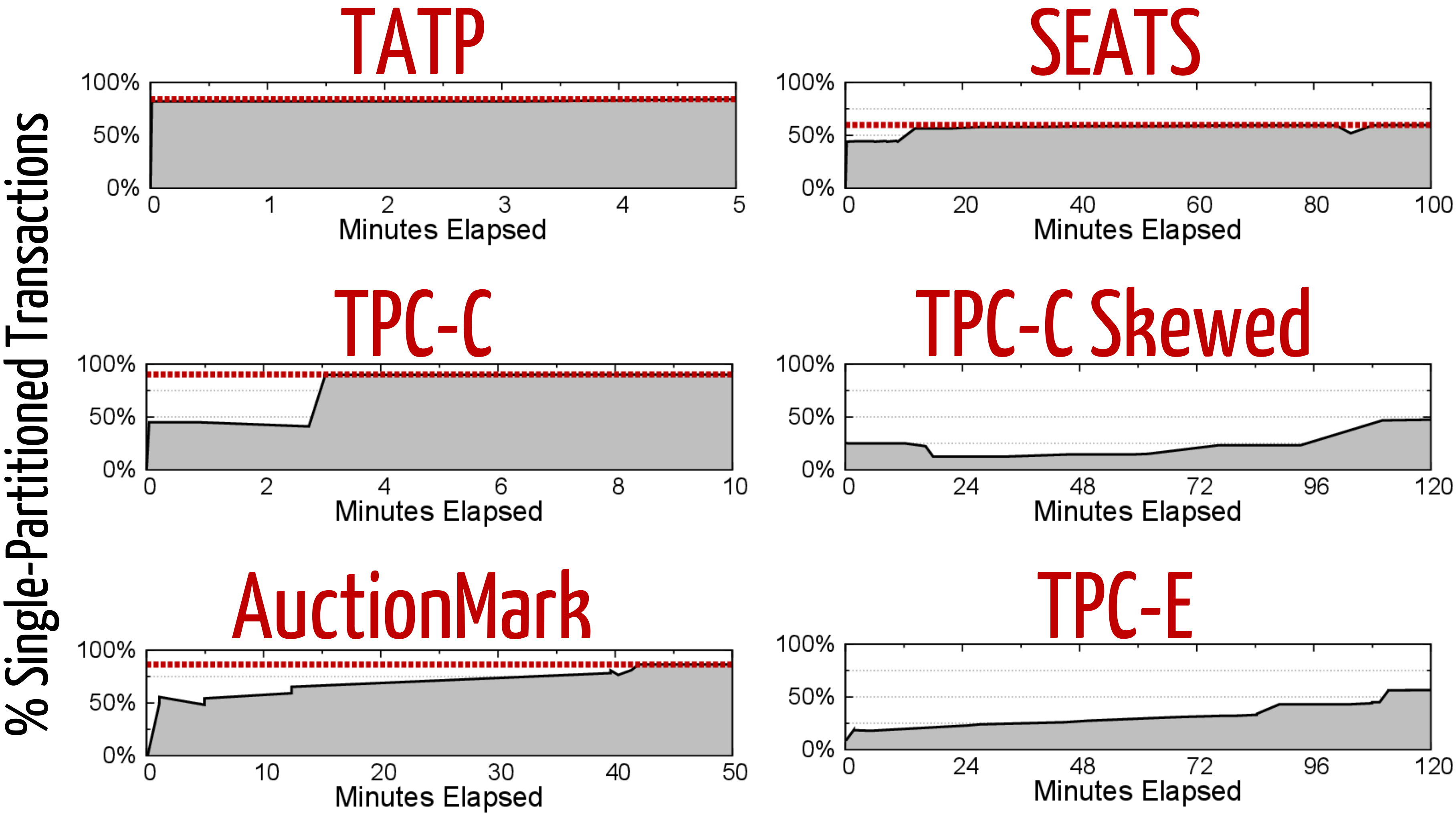


TATP
+88%

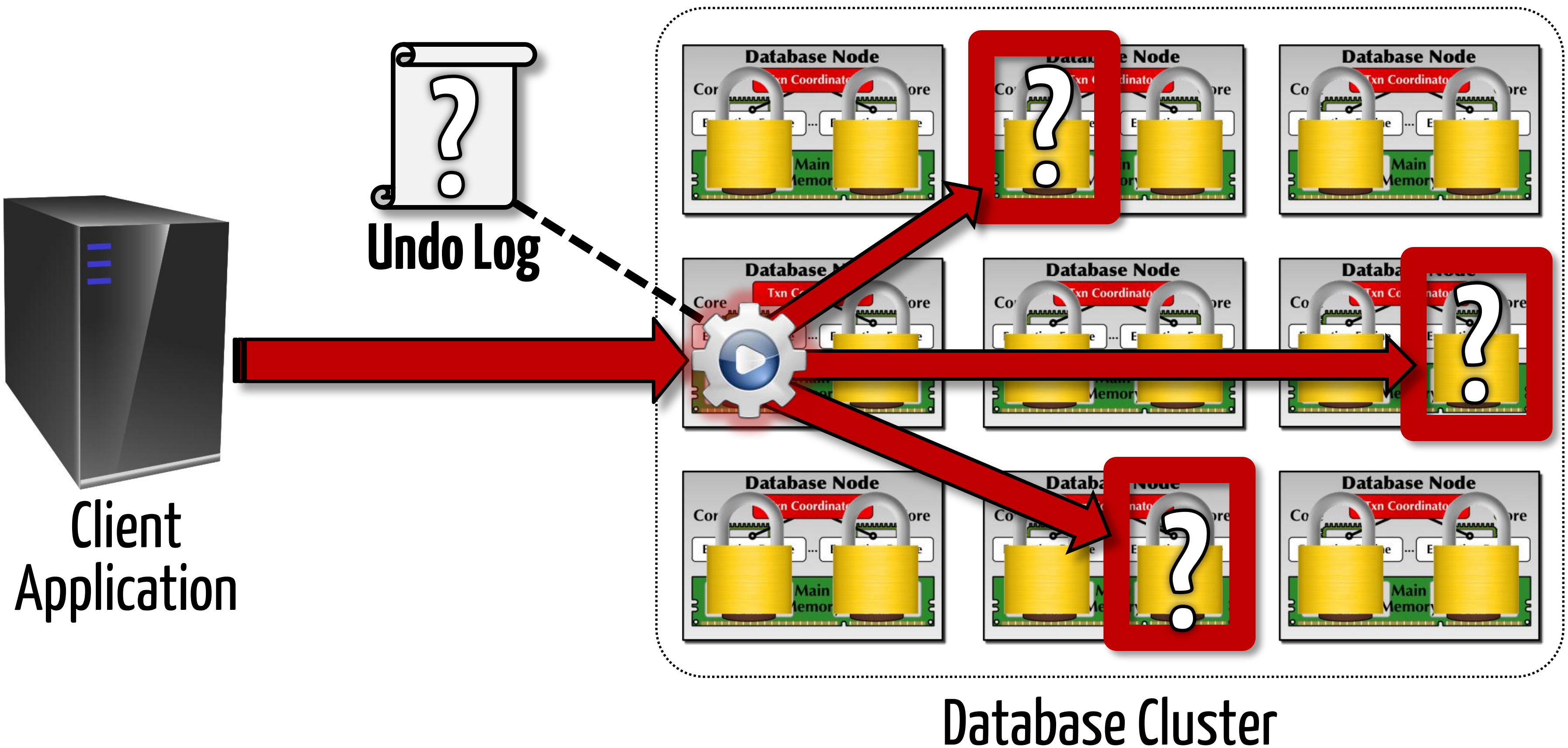
TPC-C
+16%

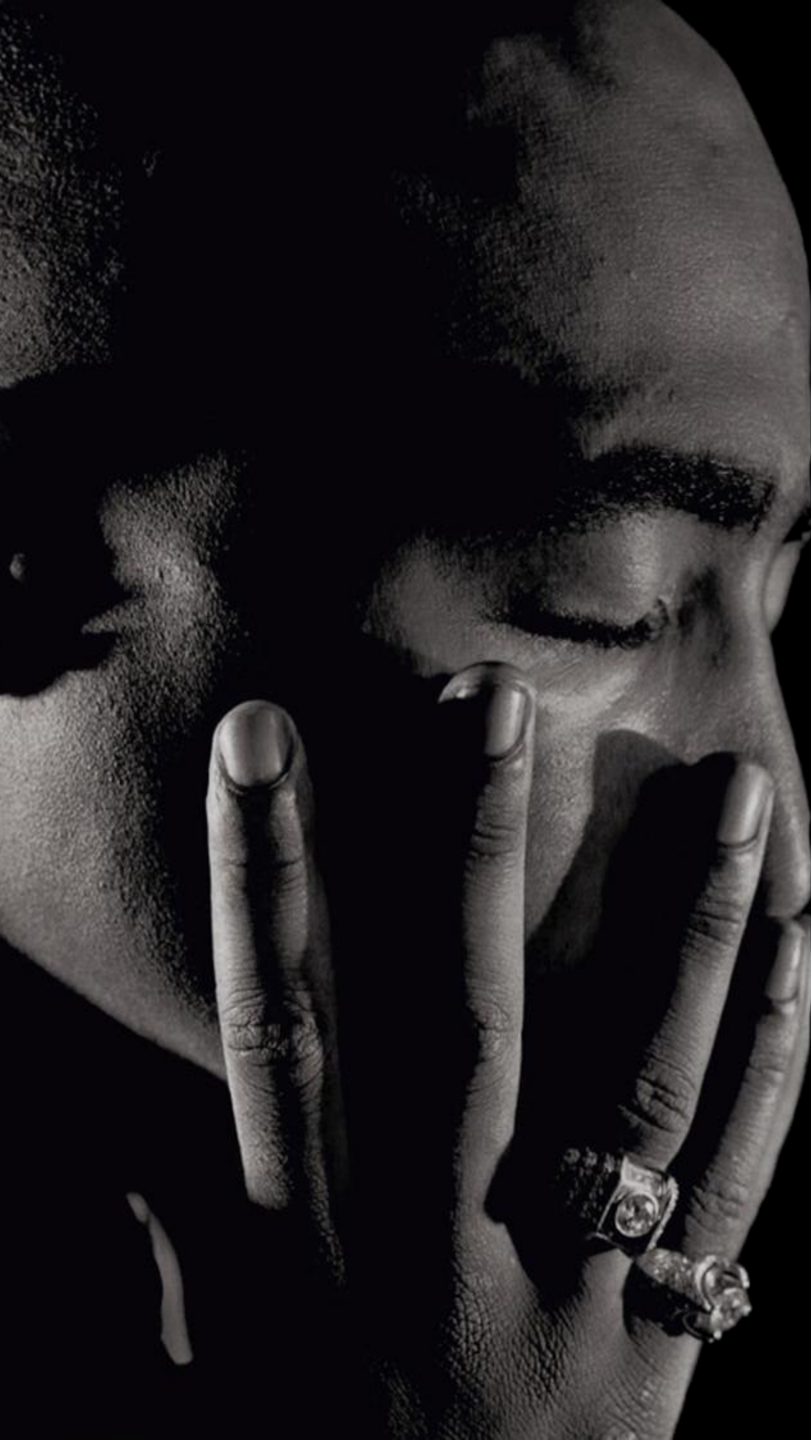
TPC-C Skewed
+183%

Search Times



H-Store


Undo Log
Client
Application
Database Node
Txn Coordinator
Main
Memory
Database Cluster



Grab your glocks
when you see 2pac.

Call the cops
when you see 2pac.

Hit 'Em Up (1996)

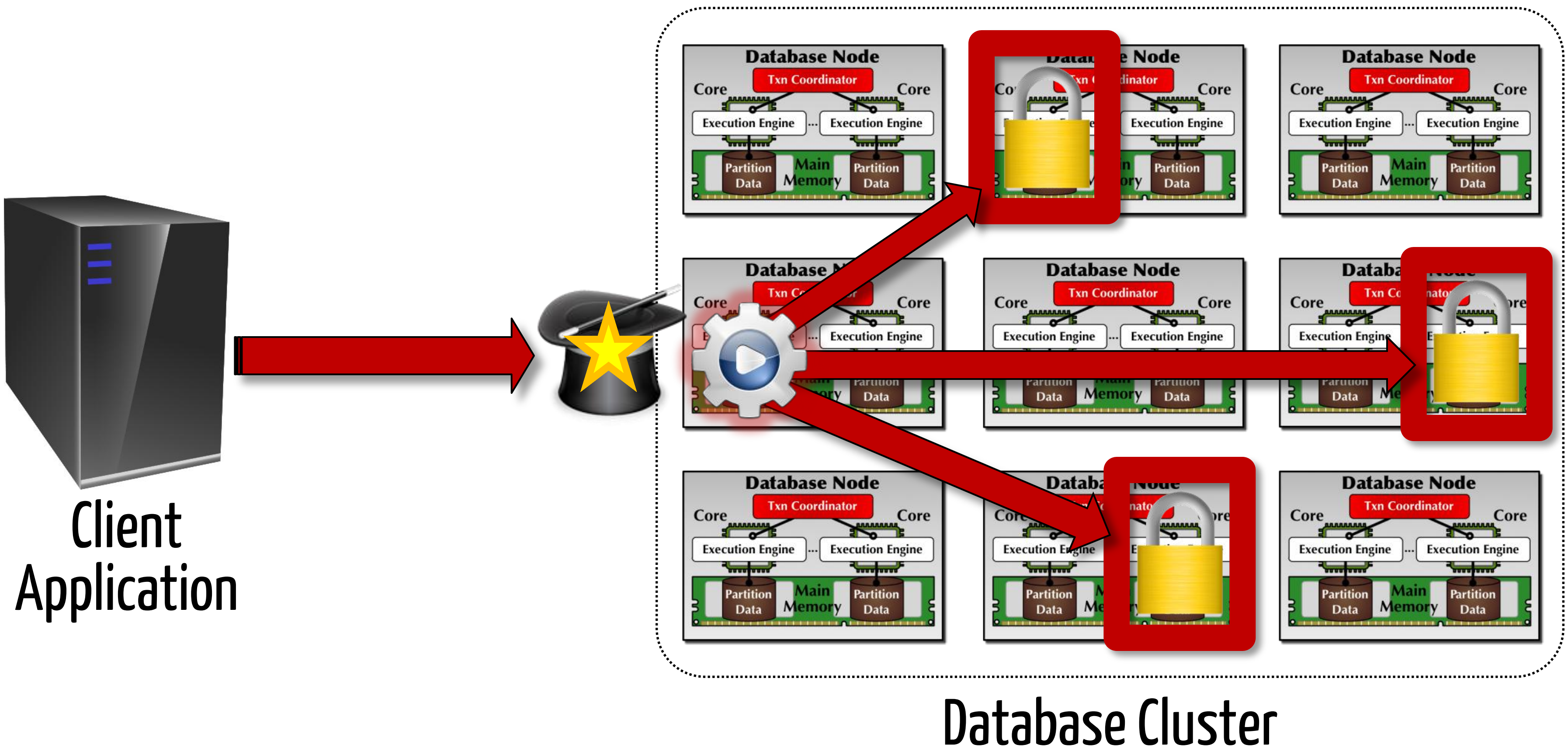
Optimization #2:

Predict what txns will do <u>before</u> they execute.

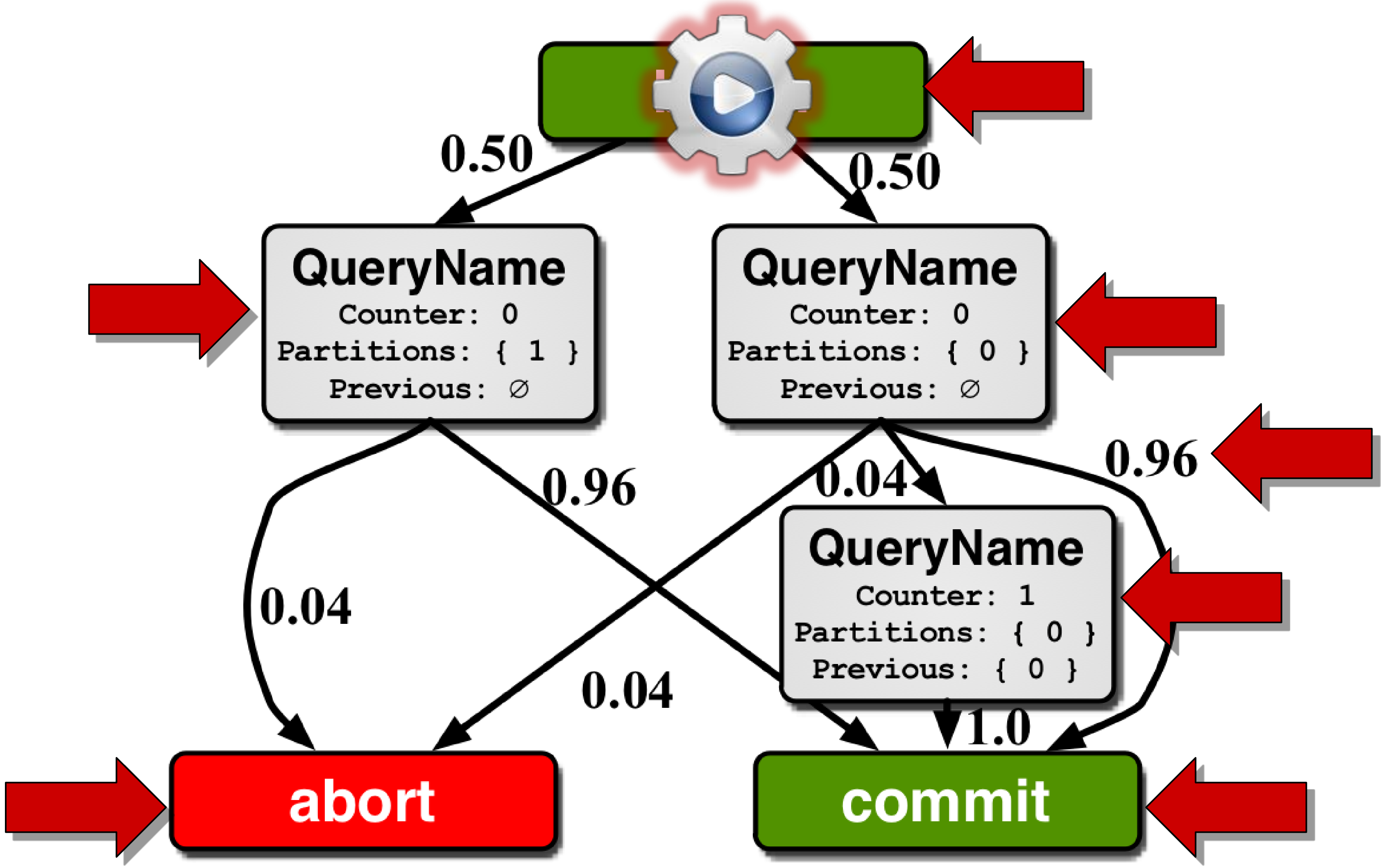
On Predictive Modeling for Optimizing Transaction Execution in Parallel OLTP Systems
Proc. VLDB Endow., vol. 5, pp. 85-96, 2011.


On Predictive Modeling for Optimizing Transaction Execution in Parallel OLTP Systems

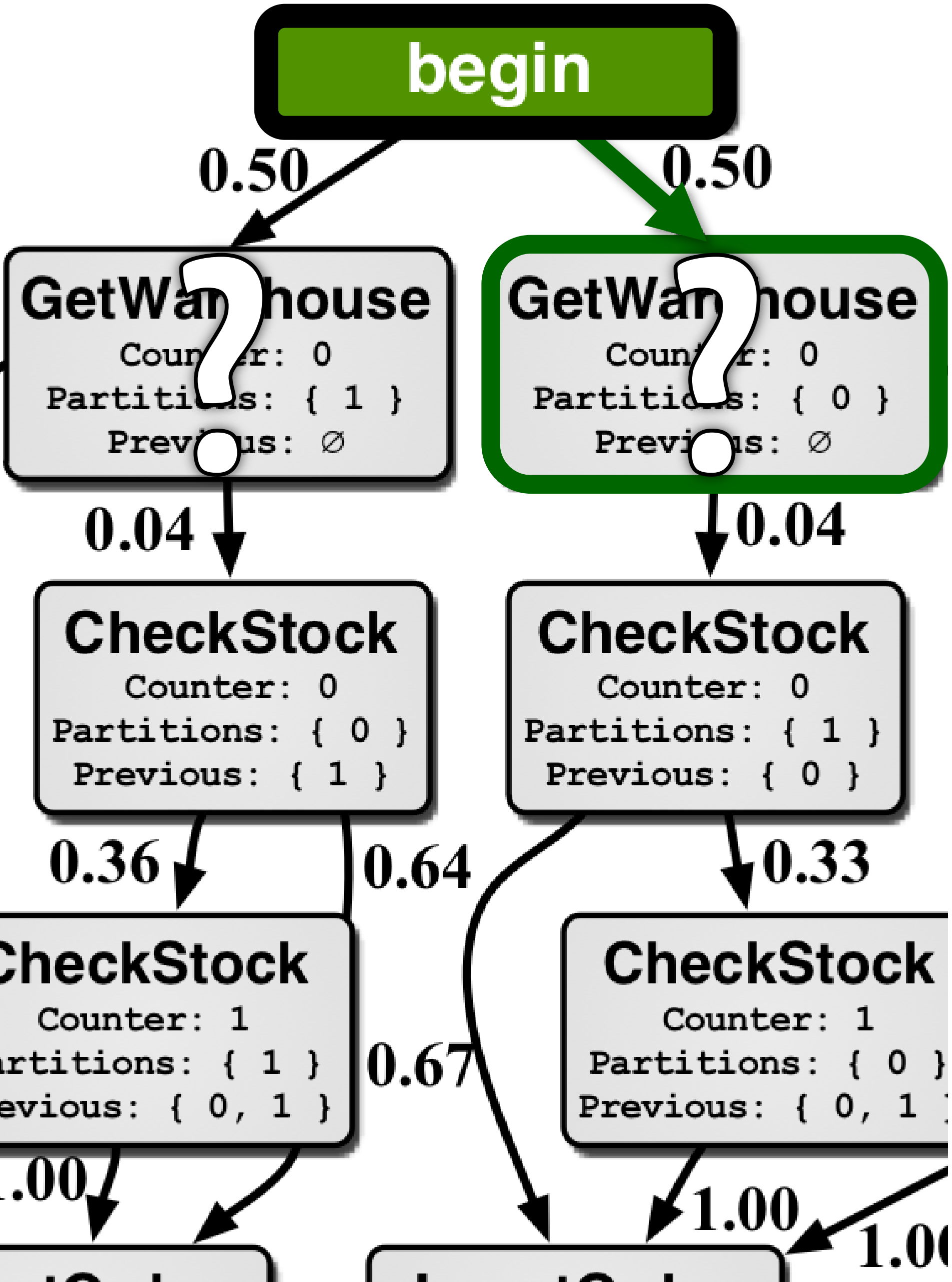

Houdini


Client
Application
Database Node
Txn Coordinator
Core
Core
Execution Engine
Execution Engine
Partition
Data
Main
Memory
Partition
Data
Database Cluster

Houdini


0.50
0.50
QueryName
Counter: 0
Partitions: { 1 }
Previous: ∅
QueryName
Counter: 0
Partitions: { 0 }
Previous: ∅
0.96
0.04
0.96
QueryName
Counter: 1
Partitions: { 0 }
Previous: { 0 }
0.04
0.04
1.0
abort
commit

Current State

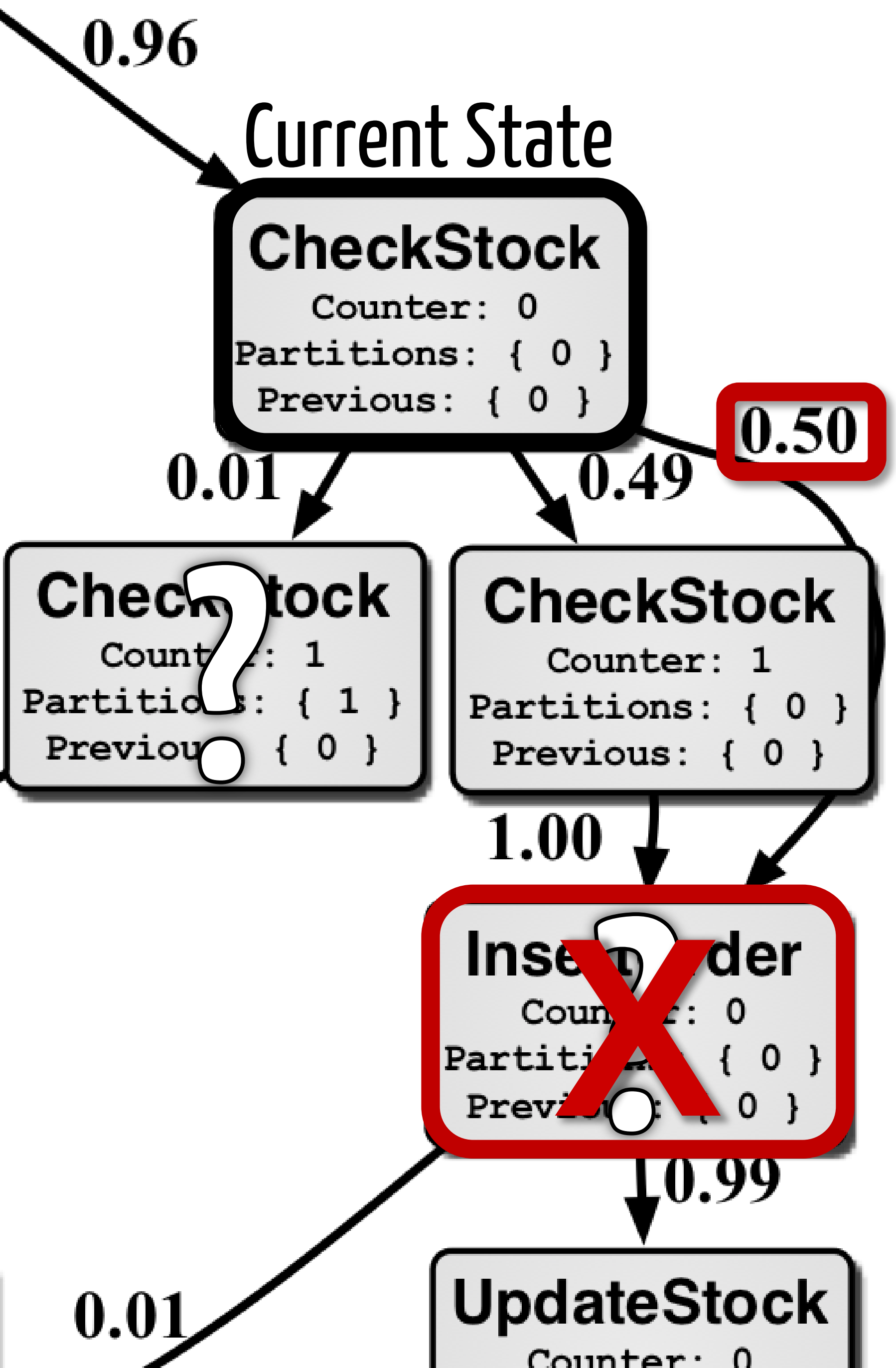


Input Parameters:

```
w_id=0
i_w_ids=[0,1]
i_ids=[1001,1002]
```

GetWarehouse:

```
SELECT * FROM WAREHOUSE
 WHERE W_ID = ?
```



Input Parameters:

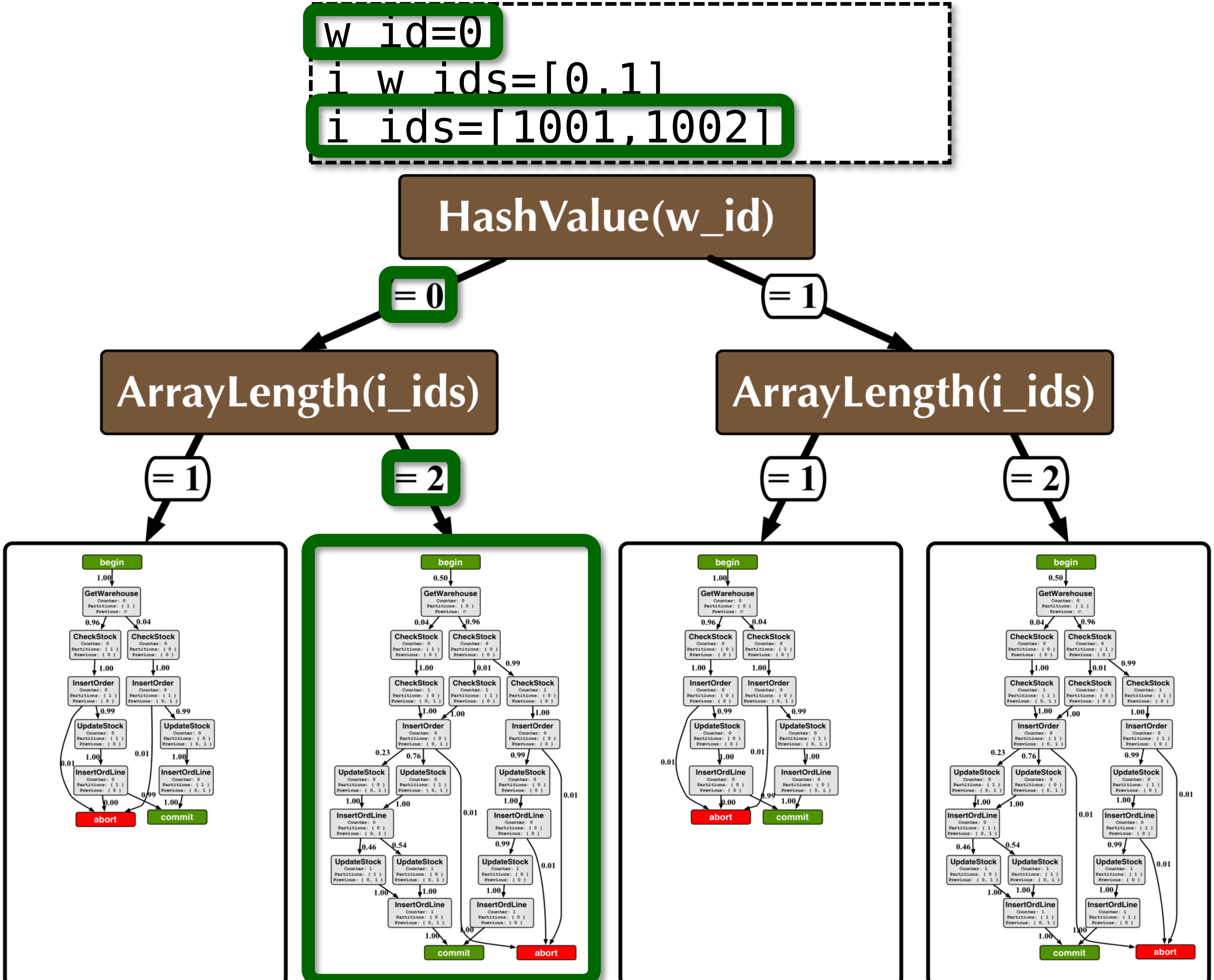
```
w_id=0
i_w_ids=[0,1]
i_ids=[1001,1002]
```

CheckStock:

```
SELECT S_QTY FROM STOCK
 WHERE S_W_ID = ?
   AND S_I_ID = ?;
```

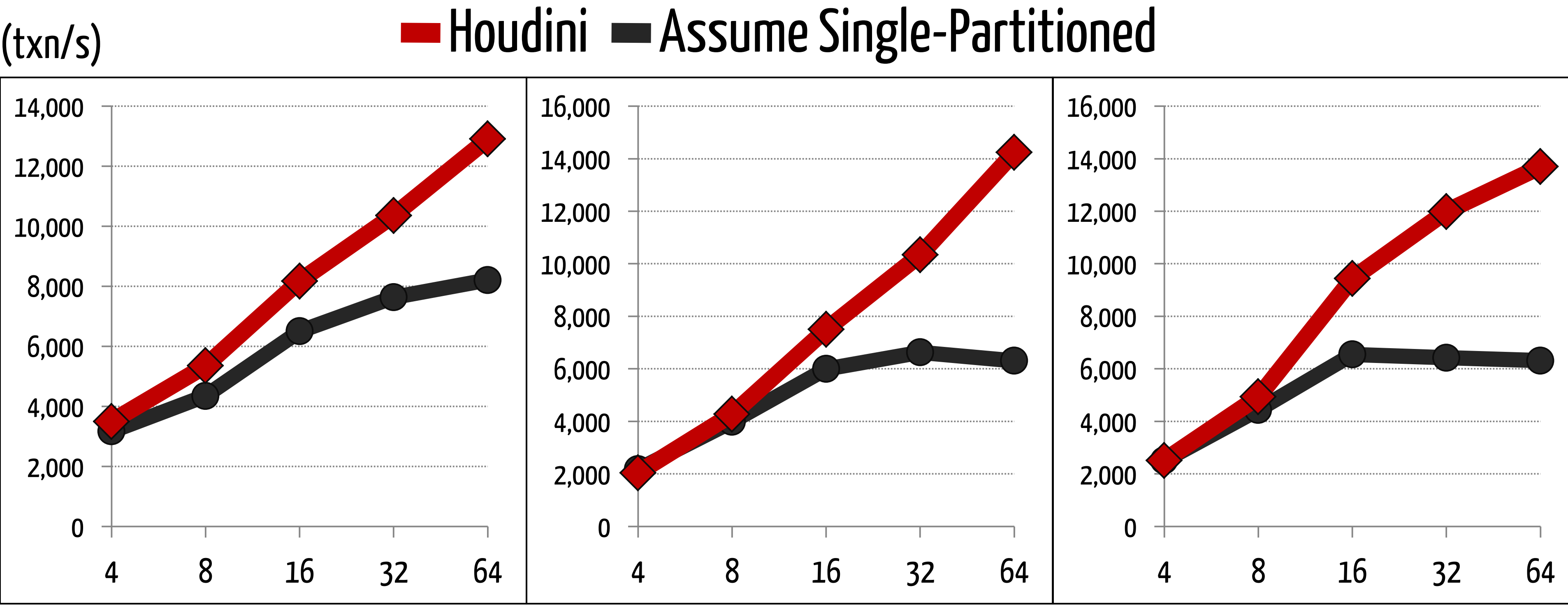
InsertOrder:

```
INSERT INTO ORDERS
(o_id, o_w_id)
VALUES (?, ?);
```



w_id=0
i_w_ids=[0,1]
i_ids=[1001,1002]
HashValue(w_id)
= 0
= 1
ArrayLength(i_ids)
ArrayLength(i_ids)
= 1
= 2
= 1
= 2

Throughput

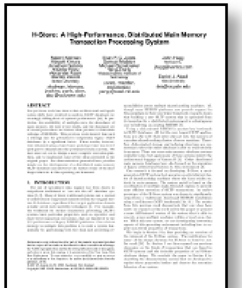


TATP **+57%**

TPC-C **+126%**

AuctionMark **+117%**

Recap



Optimize Single-Partition Execution

H-Store: A High-Performance, Distributed Main Memory Transaction Processing System
R. Kallman, H. Kimura, J. Natkins, A. Pavlo, S. Zdonik, E. P. C. Jones, S. Madden, M. Stonebraker, et al.
Proc. VLDB Endow., vol. 1, iss. 2, pp. 1496-1499, 2008.

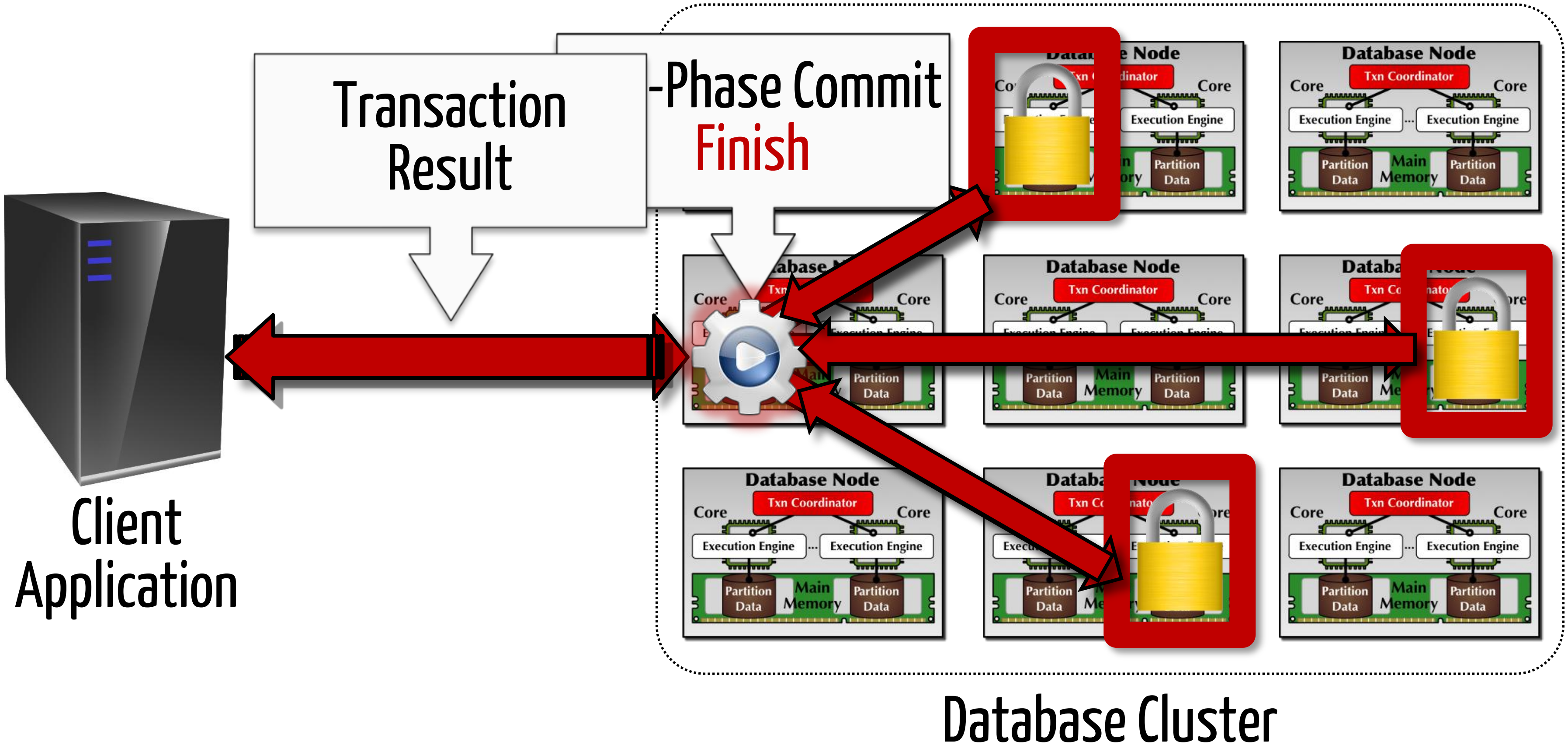
Minimize Distributed Transactions

Skew-Aware Automatic Database Partitioning in Shared-Nothing, Parallel OLTP Systems
A. Pavlo, C. Curino, and S. Zdonik,
Proceedings of SIGMOD, 2012.

Identify Distributed Transactions

On Predictive Modeling for Optimizing Transaction Execution in Parallel OLTP Systems
A. Pavlo, E. P. C. Jones, and S. Zdonik
Proc. VLDB Endow., vol. 5, pp. 85-96, 2011.

Two-Phase Commit



Thesis Proposal:

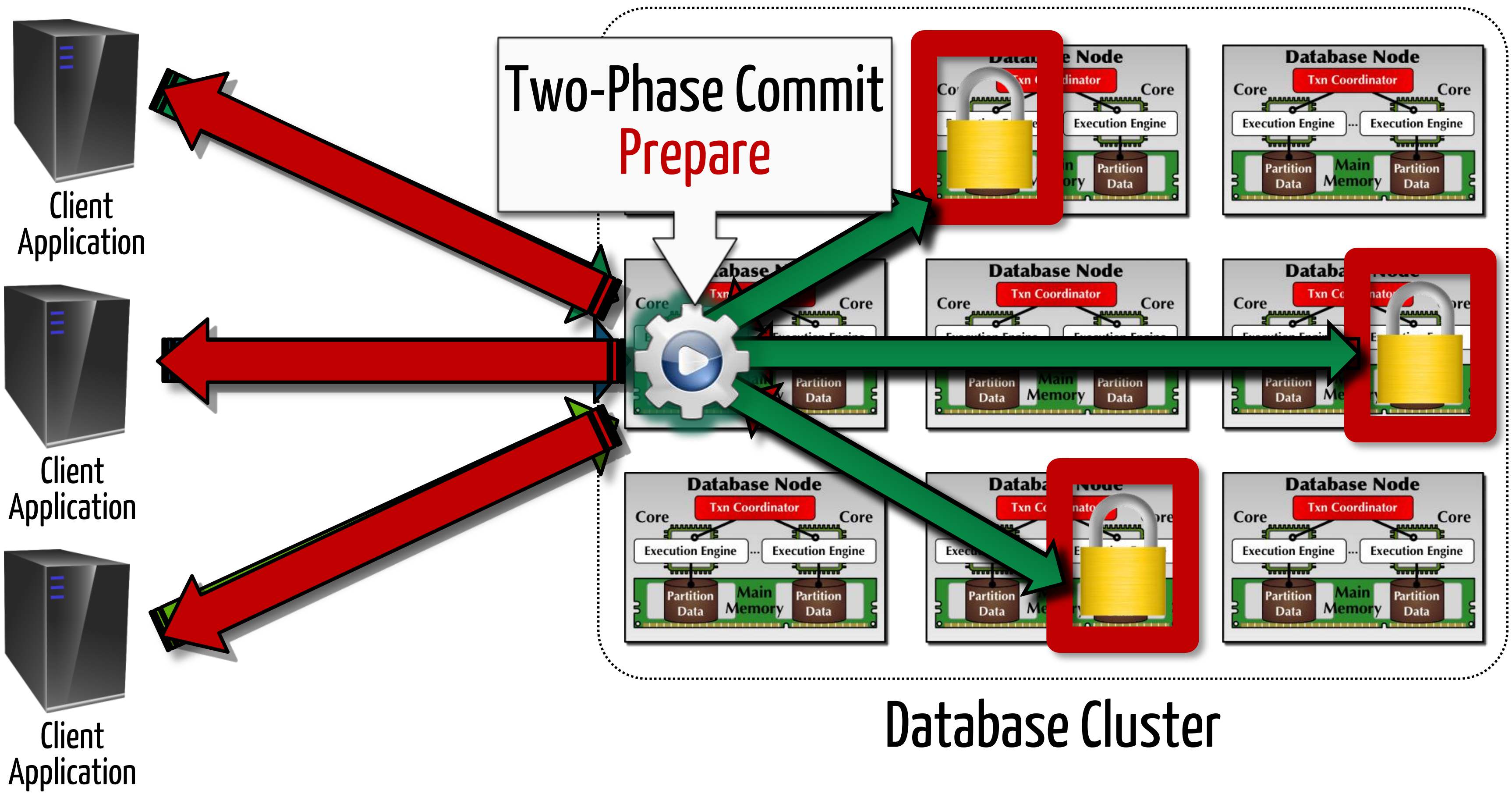
Amortize distributed txn coordination with creative scheduling.

Dr. Zdonik – Or How I Learned to Stop Worrying
And Love Distributed Transactions
Work in Progress

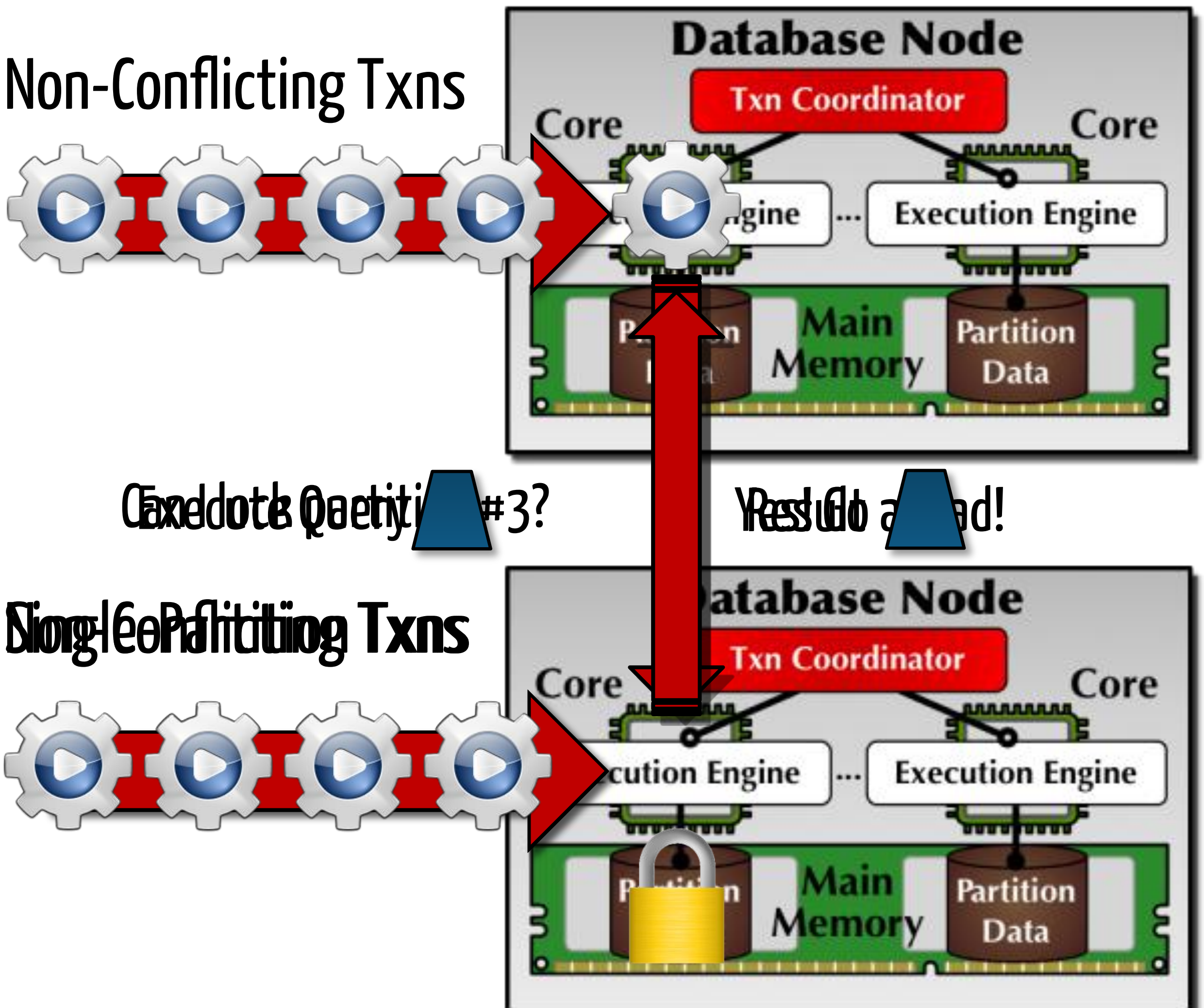

On Predictive Modeling for Optimizing Transaction Execution in Parallel OLTP Systems

ABSTRACT

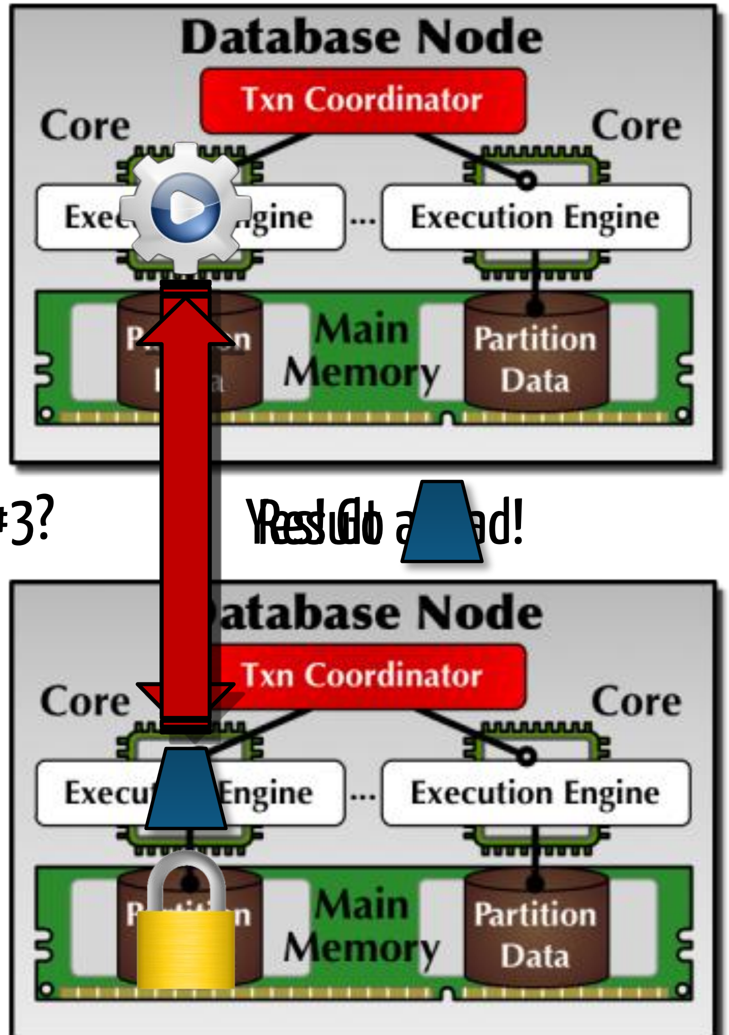
Transaction Batching



Speculative Transactions



Speculative Queries



Can I lock partition #3?

Execute Query

Yes! Go ahead!

Result

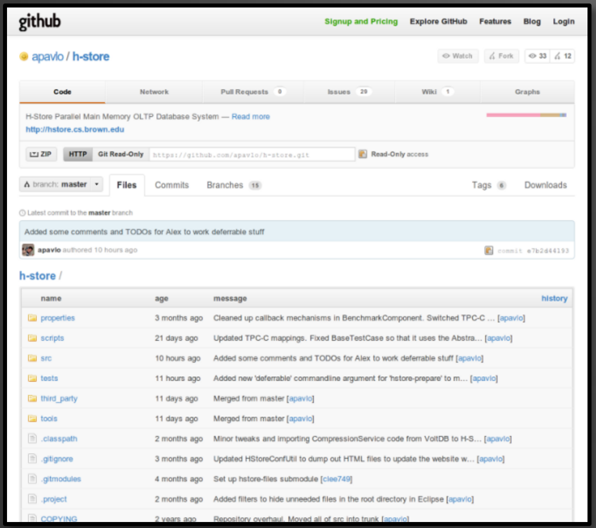
Conclusion:

Achieving fast performance is more than just using only RAM.

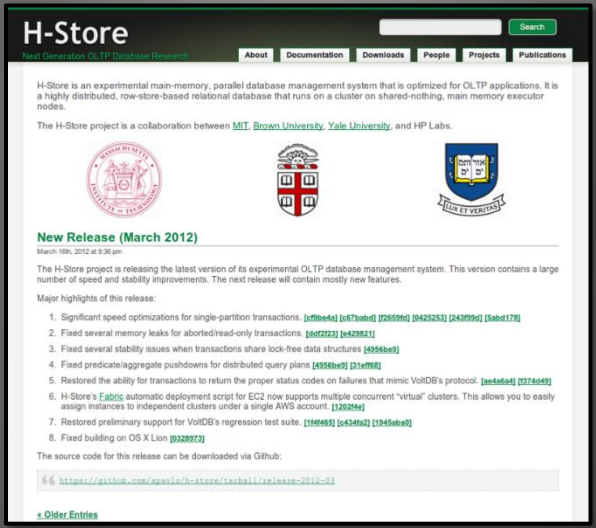
Proposed Schedule:

Coding+Experiments (Summer)
Paper Submission (Fall/Winter)
Thesis Writing (Spring)

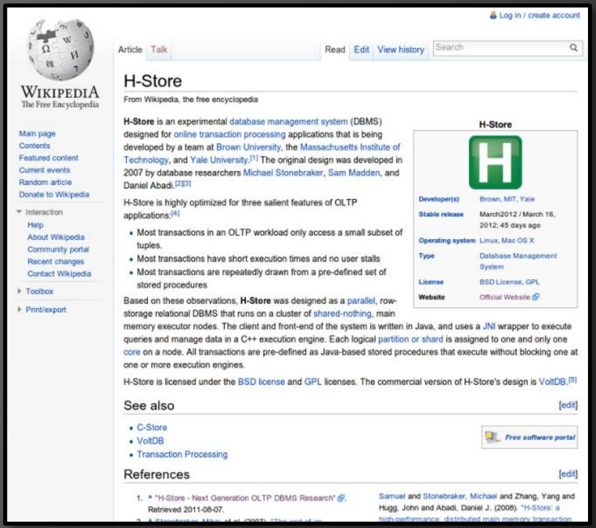
H-Store



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



http://hstore.cs.brown.edu



http://wikipedia.org/wiki/H-Store