

# H-STORE

## A High-Performance, Distributed Main Memory Transaction Processing System



Robert Kallman  
Jonathan Natkins  
Hideaki Kimura

Andrew Pavlo  
Alexander Rasin  
Stanley Zdonik



Evan P. C. Jones  
Samuel Madden  
Michael McCanna

Michael Stonebraker  
Yang Zhang



Ryan Betts  
Bobbi Heath

John Hugg



Daniel J. Abadi

### Overview

#### » Many OLTP databases have common properties:

- Repetitive execution of short-lived transactions.
- Entire database fits in main memory of a cluster.
- Data naturally partitions on keys (e.g., customer ids, warehouse ids)
- Individual transaction invocations only touch a small subset of data.

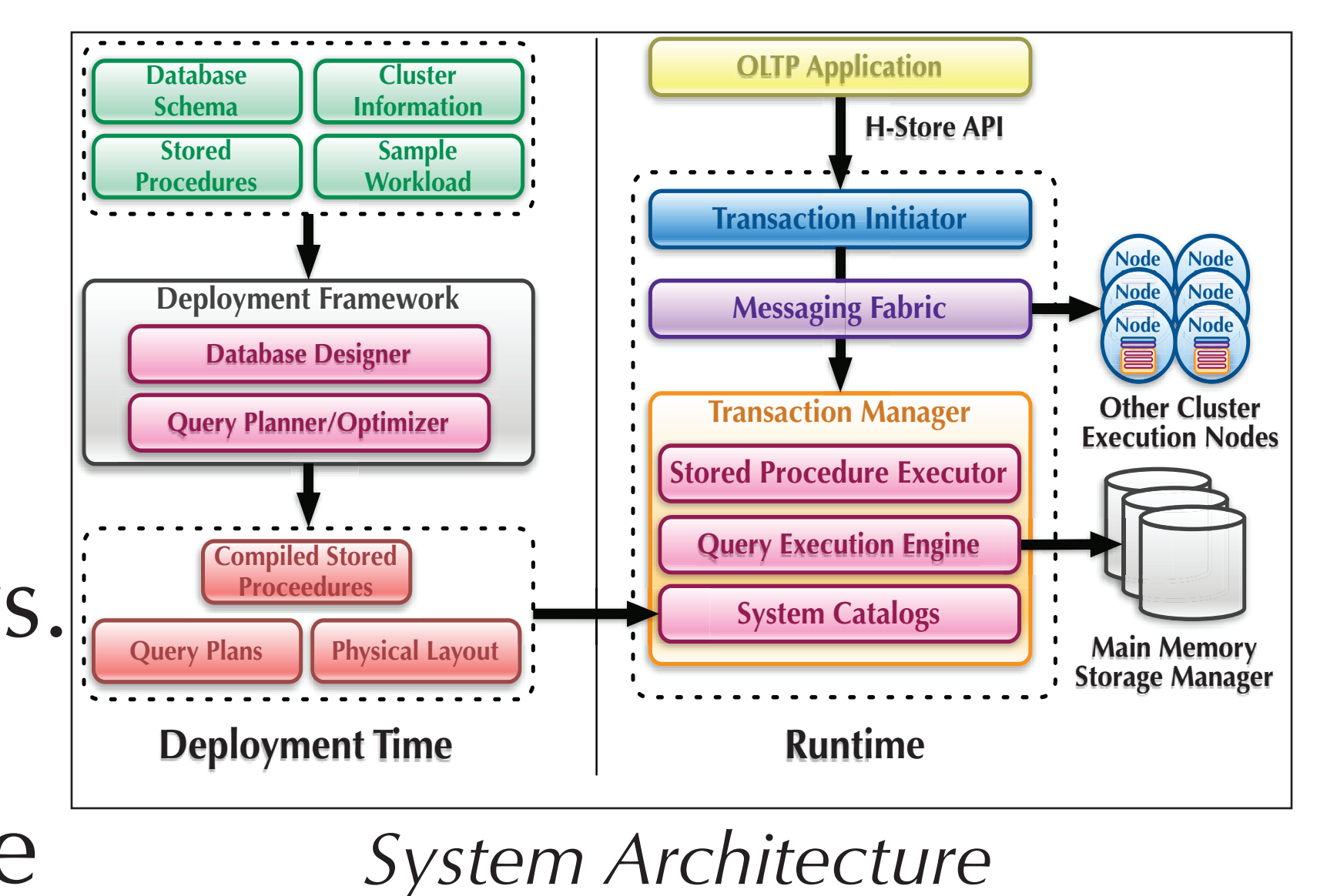
#### » The development of H-Store facilitates research into exploiting these non-trivial aspects of OLTP systems.

#### » Previous main-memory databases have focused on the migration of legacy features from disk-based environments.

### System Design

#### » H-Store is a new clean-slate design:

- Main-memory row storage.
- Designed for multi-core machines.
- One execution thread per database partition.
- Distributed operation on shared-nothing clusters.
- Applications invoke pre-defined stored procedures consisting of structured control code intermixed with parameterized SQL commands.



#### » Replication provides data durability:

- Local replication for quick fail over.
- Remote replication for disaster recovery.

### OLTP Transactions

#### » H-Store's transaction protocol is optimized for fast single-sited transactions:

- Such transactions are able to execute to completion without retrieving intermediate data from other nodes.
- Serial execution eliminates the need for concurrency control mechanisms.
- An automatic database designer will aid in the deploying databases with configuration that maximizes the number of single-sited transactions.

#### » Complex multi-site transactions are supported, but require more heavy-handed concurrency protocols.

### Not Just Another Main Memory RDBMS

- » Uses data partitioning to distribute execution on shared-nothing, multi-core clusters
- » No disk-based logging or locking
- » Current prototype is 4x faster than a well-known commercial database

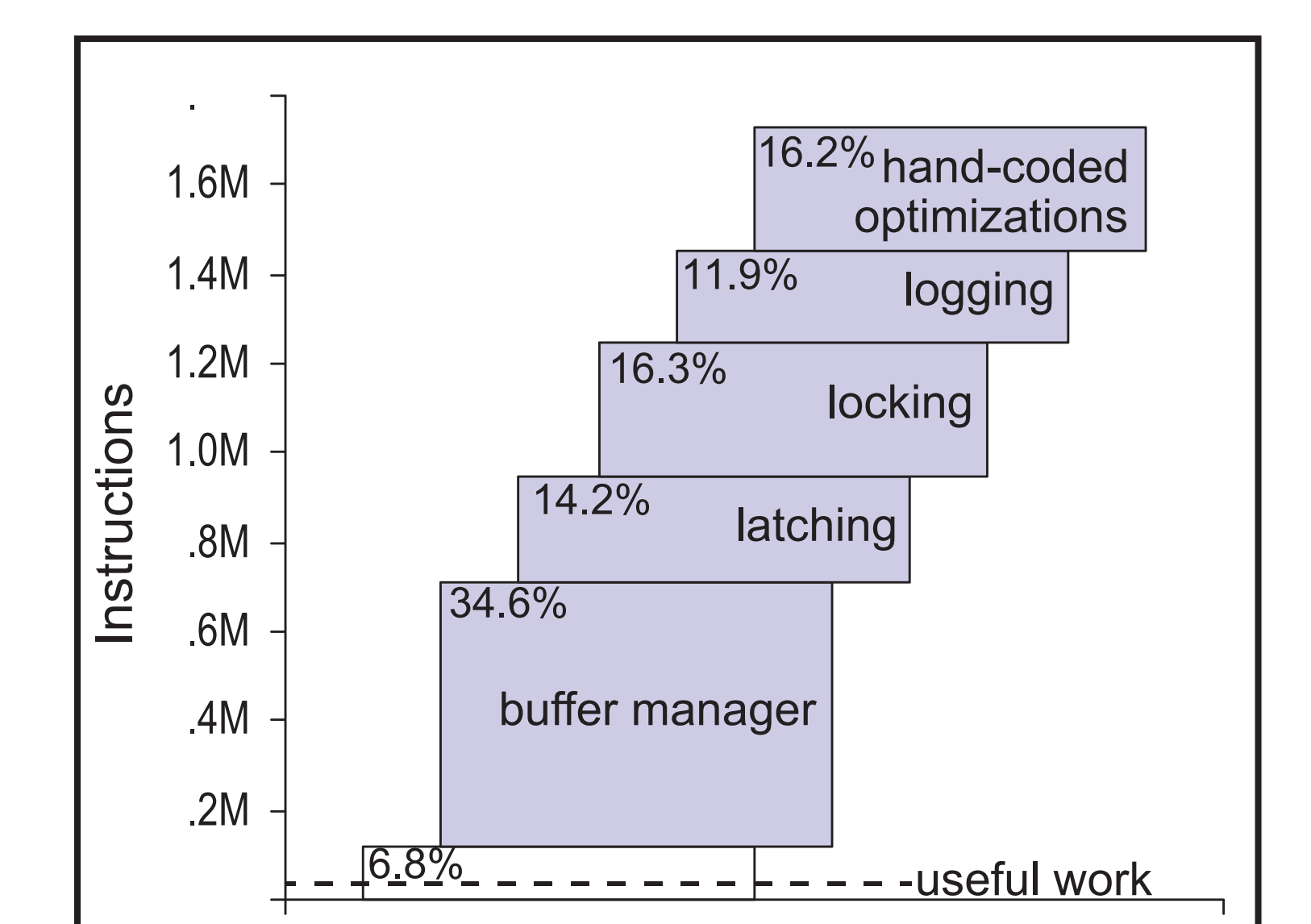
### Performance

#### » Our previous work shows that a specialized database engines can outperform "one-size fits all" systems:

- A traditional row-storage RDBMS was shown to perform little useful work during OLTP workload experiments.

#### » H-Store TPC-C performance numbers:

- Current Prototype: 4875 txn/s
- Commercial Database: 1207 txn/s
- Tested on a 2-core Intel Xeon E5320 @ 1.86 Ghz with 10 warehouses.



Main Memory Store Performance