H-Store

KARTHICK VENKATESAN[1,*,+]

[1] School of Informatics and Computing, Bloomington, IN 47408, U.S.A.
[*] Corresponding authors: vkarthickprabu@gmail.com
[+] HID - S17-IO-3023

paper1, March 26, 2017

"NewSQL" denotes a class of modern (RDBMS)relational database management systems that seek to provide the same scalable performance of NoSQL systems for online transaction processing (OLTP) read-write workloads while still maintaining the ACID guarantees of a traditional database system. H-Store is one the first NewSQL system which is a complete redesign of the traditional RDMS database.

Keywords: H-Store,OLTP, I524

https://github.com/karvenka/sp17-i524/tree/master/paper1/S17-IO-3023/report.pdf

1. INTRODUCTION

H-Store is an experimental database management system (DBMS) designed for online transaction processing applications that was developed by a team of researchers at Brown University, Carnegie Mellon University, the Massachusetts Institute of Technology, and Yale University.

Traditional SQL databases offer rich transaction capabilities, ad-hoc query and reporting capability, and vast amounts of standards-based tooling. Where they fall short is in the ability to scale out to meet the needs of modern, high-performance applications.NoSQL Databases are a solution to issues of scale that emerged as organizations began dealing with immense volumes, velocity, and variety of big data from a multitude of sources. NoSQL solutions offered availability, flexibility and scale. However, they came with difficult tradeoffs - sacrifice of data consistency and transactional (ACID) guarantees, loss of ad-hoc query capability and increased application complexity.By removing just the parts of SQL that hampered scalability and performance, it was able to increase both while maintaining key advantages of SQL systems.NewSQL technologies offer the best of both worlds: the scale of NoSQL with the ACID guarantees, strong consistency, minimized application complexity and interactivity of SQL. The H-Store DBMS belongs to this class of parallel database management systems, called NewSQL.

The current RDBMS systems were architected more than 25 years ago, when hardware characteristics were much different than today. Today processors are thousands of times faster and memories are thousands of times larger. Disk volumes have increased enormously, making it possible to keep essentially everything, if one chooses to in memory [1]. The current day DBMS include the following architectural features which are not fully utilising the technological advancements in the last 25 years .

- Disk oriented storage and indexing structures
- Multithreading to hide latency
- Locking-based concurrency control mechanisms
- Log-based recovery

H-Store is a complete redesign of the traditional RDMS database designed to fully leverage the advancements in hardware capabilities over the last 25 years and experimetal results show that the OLTP processing on a H-Store database is faster by a factor of 82 [1].

2. SYSTEM ARCHITECTURE

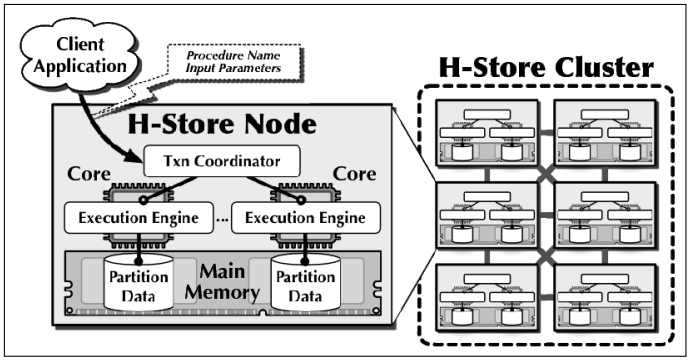

Fig. 1. H-Store Architecture [2]

A single H-Store instance as in Figure 1 consists of a cluster of two or more computational nodes deployed within the same domain [2].

- A node is a single physical computer system that hosts one or more execution sites and one transaction coordinator.
- A site is the logical operational entity in the system; it is a single-threaded execution engine that manages some portion of the database and is responsible for executing transactions on behalf of its transaction coordinator.
- The transaction coordinator is responsible for ensuring the serializability of transactions along with the coordinators located at other nodes.
- A typical H-Store node contains one or more multi-core CPUs, each with multiple hardware threads per core. As such, multiple sites are assigned to nodes independently and do not share any data structures or memory.

3. STORAGE ARCHITECTURE

Every table in H-Store is horizontally divided into multiple shards that each consist of a disjoint sub-section of the entire table. The boundaries of a table's fragments are based on the selection of a column for that table; each tuple is assigned to a fragment based on the hash values of this attribute. H-Store also supports partitioning tables on multiple columns. Related fragments from multiple tables are combined together into a partition that is distinct from all other partitions. Each partition is assigned to exactly one site.

All tuples in H-Store are stored in main memory on each node; the system never needs to access a disk in order to execute a query.It replicates partitions to ensure both data durability and availability in the event of a node failure. Data replication in H-Store occurs in two ways: (1) replicating partitions on multiple nodes and (2) replicating an entire table in all partitions. For the former, H-Store adopts the k-safety concept, where k is defined by the administrator as the number of node failures a database can tolerate before it is deemed unavailable [2].

4. TRANSACTION PROCESSING

Client applications make calls to the H-Store system to execute pre-defined stored procedures. Each procedure is identified by a unique name and consists of user-written Java control code intermixed with parameterized SQL commands. The input parameters to the stored procedure can be scalar or array values, and queries can be executed multiple times.

Each instance of a stored procedure executing in response to a client request is called a transaction. Similarly as with tables, stored procedures in H-Store are assigned a partitioning attribute of one or more input parameters. When a new request arrives at a node, the transaction coordinator hashes the value of the procedure's partitioning parameter and routes the transaction request to the site with the same id as the hash. Once the request arrives at the proper site, the system invokes the procedure's Java control code, which will use the H-Store API to queue one or more queries in a batch. The control code invokes another command in H-Store to block the transaction while the execution engine executes the batched queries.

All of operations within a transaction are atomic across all the partitions involved in the transaction. When a transaction executes, it has exclusive access to the data at the partitions that it locks, therefore all of its operations execute on a consistent view of the database and are isolated from all other transactions (since no other transactions execute concurrently at those partitions). Finally, with snapshots and command logging, the state of the database is completely recoverable and all changes made by transactions are durable [2] .

5. DOCUMENTATION

- Detailed documentation on H-Store Deployment , Configuration,Debugging and Development is available at [3]
- H-Store DBMS is open source and complete source code is available in the [4] github repository.

6. COMMERCIAL USE

A commercial implementation of H-Store is VoltDB. VoltDB was developed for production environments, and thus it is focused on high-performance throughput for single-partition transactions and provides robust handling of failures that are needed for a main memory system.H-Store uses VoltDB's VoltProcedure API, so any stored procedures written for VoltDB work work in H-Store [5].

7. RELATED TECHNOLOGIES-OTHER IN MEMORY NEW SQL DATABASES

- **VoltDB**:VoltDB is an in-memory, scale-out SQL database purpose-built to power a new generation of applications that thrive on fast, smart data. Tapping the lightning speed and real-time analytics of VoltDB, organizations are able to add context and intelligence to data – the instant it arrives – to make real-time transactional decisions that maximize business value [6].
- **Clustrix**:Clustrix provides the leading scale-out SQL database engineered for the cloud and the database built specially to meet the unique growth, performance and availability demands of today's e-commerce sites. ClustrixDB, helps to build business critical applications that support massive transactional volume and realtime reporting of business performance metrics. ClustrixDB delivers more than one trillion transactions per month for customers [7].
- **NuoDB**:NuoDB is a scale-out SQL database for the cloud and the modern datacenter. It is the NewSQL solution designed to meet the needs of cloud-scale apps and over 50 billion connected devices in use worldwide [8].
- **MemSQL**:MemSQL is a leader in real-time databases for transactions and analytics.As a purpose built database for instant access to real-time and historical data, MemSQL uses a familiar SQL interface and a horizontally scalable distributed architecture that runs on commodity hardware or in the cloud [9].

8. COMPARISON OF NEW SQL DATABASES

A comparison of other NewSQL database features with H-Store is provided in Fig 2.

9. USE CASE

In Memory NewSQL databases such as H-Store are ideally suited for handling uses cases related to Fast Data [11].

	Year Released	Main Memory Storage	Partitioning	Concurrency Control	Replication	Summary
Clustrix [6]	2006	No	Yes	MVCC+2PL	Strong+Passive	MySQL-compatible DBMS that supports shared-nothing, distributed execution.
CockroachDB[7]	2014	No	Yes	MVCC	Strong+Passive	Built on top of distributed key/value store. Uses softwares hybrid clocks for WAN replication.
Google Spanner[24]	2012	No	Yes	MVCC+2PL	Strong+Passive	WAN-replicated, shared-nothing DBMS that uses special hardware for timestamp generation.
H-Store[8]	2007	Yes	Yes	TO	Strong+Passive	Single-threaded execution engines per partition. Optimized for stored procedures.
Hy-Per[9]	2010	Yes	Yes	MVCC	Strong+Passive	HTAP DBMS that uses query compilation aand memory eficiente indexes.
MemSQL[11]	2012	Yes	Yes	MVCC	Strong+Passive	Distributed, shared-nothing DBMS using compiled queries. Supports MySQL wire protocol.
NuoDB[14]	2013	Yes	Yes	MVCC	Strong+Passive	Split architecture with multiple in-memory executor nodes and a single shared storage node.
SAPHANA[55]	2010	Yes	Yes	MVCC	Strong+Passive	Hybrid storage(rows + cols).Amalgamation of previus TREX, P*TIME, and MaxDB systems.
VoltDB[17]	2008	Yes	Yes	TO	Strong+Active	Single-threaded execution engines per partition.Supports streaming operators.

Fig. 2. NewSQL DB Comparison

[10]

- Real-time recommendation engines or hyper-personalization applications that can detect and act on individual customer needs in real-time.Emagine International is a innovative company provides a real-time, adaptive contextual marketing platform and managed marketing services for mobile service providers. To differentiate its offering and help operators take advantage of real-time analytics and decisioning, Emagine implemented H-Store's Commercial implementation VoltDB as the core of its Emagine Real-time Event Decisioning (ERED) platform. This has allowed Emagine to architect a fast data solution that requires 3 milliseconds for the ingest-analyze-decide journey through the ERED platform, enabling Emagine to deliver customized offers to subscribers in fewer than 250 milliseconds. Mobile operators that use the ERED platform achieved a measurable ROI in real-time personalization [12].

- Real-time, "down to the last dollar" resource management applications. More than 150,000 apps and the world's leading advertisers rely on Airpush a mobile advertising company to deliver the industry's highest performance, driven by exceptional ad formats and targeting technology.The company originally built a MySQL infrastructure to manage advertising transactions. Airpush's business grew exponentially, pushing the limits of MySQL. To ensure the system would be able to process transactions fast enough to align advertising purchases with customer budgets, Airpush began to look into alternative. Airpush selected VoltDB (H-Store commercial implementation), which offererred the performance of in-memory, the scalability of NoSQL, and the transactional consistency of traditional relational databases. With VoltDB, Airpush built a transactional, database oriented applications against data feeds that were previously limited to stream processing methods because of scale. Supporting hundreds of thousands of concurrent connections with round-trip latencies in milliseconds, VoltDB is an ideal platform for high-speed policy enforcement, authorization, rule evaluation, and quota management [13].

10. CONCLUSION

H-Store and its commercial implementation VoltDB are main memory DBMS designed for applications running on a cluster of nodes needing high transaction throughput. They are ideally suited for the Fast Data use case of Big Data. As an example of a "NewSQL" DBMS with superior performance, they are ideally positioned to take over the Modern Transaction Processing market from legacy disk-oriented RDBMS.

ACKNOWLEDGEMENTS

The authors thank Prof. Gregor von Laszewski for his technical guidance.

REFERENCES

[1] M. Stonebraker, S. Madden, D. J. Abadi, S. Harizopoulos, N. Hachem, and P. Helland, "The end of an architectural era: (it's time for a complete rewrite)," in *VLDB '07: Proceedings of the 33rd international conference on Very large data bases*. VLDB Endowment, 2007, pp. 1150–1160. [Online]. Available: http://hstore.cs.brown.edu/papers/hstore-endofera.pdf

[2] Brown University and Carnegie Mellon University and Massachusetts Institute of Technology and Yale University, "H-Store-Arch," Web Page, accessed 2017-02-19. [Online]. Available: http://hstore.cs.brown.edu/documentation/architecture-overview/

[3] ——, "H-store documentation," Web Page, Dec. 2012, accessed 2017-02-19. [Online]. Available: http://hstore.cs.brown.edu/documentation/

[4] ——, "H-store," Code Repository, Sep. 2010, accessed: 2017-2-19. [Online]. Available: https://github.com/apavlo/h-store

[5] ——, "H-Store-Faq," Web Page, accessed 2017-02-19. [Online]. Available: http://hstore.cs.brown.edu/documentation/faq

[6] VoltDB, Inc, "VoltDB," Web Page, accessed 2017-02-19. [Online]. Available: http://voltdb.com/

[7] Clustrix, "ClustrixDB," Web Page, accessed 2017-02-19. [Online]. Available: http://www.clustrix.com/summary-of-our-db/

[8] NuoDB, Inc, "NuoDB," Web Page, accessed 2017-02-19. [Online]. Available: https://www.nuodb.com/product-overview

[9] MemSQL Inc, "MemSQL," Web Page, accessed 2017-02-19. [Online]. Available: http://www.memsql.com/product/

[10] A. Pavlo and M. Aslett, "What's really new with newsql?" *SIGMOD Rec.*, vol. 45, no. 2, pp. 45–55, Sep 2016. [Online]. Available: http://db.cs.cmu.edu/papers/2016/pavlo-newsql-sigmodrec2016.pdf

[11] TechTarget, "fast data," Web Page, accessed 2017-02-19. [Online]. Available: http://whatis.techtarget.com/definition/fast-data

[12] VoltDB, Inc, "Emagine implements realtime telecommunications personalization with fast data," VoltDB, Inc, Case Study, 2015, accessed 2017-02-19. [Online]. Available: https://www.voltdb.com/hubfs/content/case_studies/hv_case_studies/hv-case-study-emagine-implements-real-time-telecommunications-presonalization-with-fast-data.pdf

[13] ——, "Airpush," VoltDB, Inc, Case Study, 2015, accessed 2017-02-19. [Online]. Available: https://www.voltdb.com/hubfs/content/case_studies/hv_case_studies/hv-case-study-airpush-on-a-mission-to-redefine-mobile-advertising.pdf