



memsql

NYC Data Business



---

## REAL-TIME INFRASTRUCTURE

December 18, 2012

## FOUNDERS



**Eric Frenkiel, CEO**

**facebook**



**Nikita Shamgunov, CTO**

**facebook**



## CUSTOMERS



## COMPANY

Founded: 2011

15 employees

Based in San Francisco, New York

## INVESTORS

NEA



Paul Buccheit, creator of Gmail

Aaron Levie, CEO of Box

Max Levchin, cofounder of Paypal

- Big Data is only useful if it's accessible
- SQL is a must-have for analytics
- Scale must be achieved on commodity hardware





**PRODUCT.**

MemSQL is an

in-memory, distributed, multi-tenant, and high-performance SQL database that runs on top of Apache Cassandra.





- Time value of data
- Examples: clicks, tick data, machine data, impressions
- Difficult to query real-time *and* historical data simultaneously



- MemSQL can consume millions of events/second
- Issue concurrent OLAP queries via SQL interface against lock-free data structures
- Flexible and interoperable with existing tools
- Sub-millisecond response times on terabyte-scale workloads



## ■ The tech giants

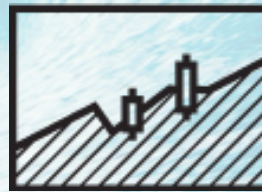
-  – BigTable, Dremel, LevelDB
-  – SCUBA
-  – Storm
-  – MemSQL

## ■ The industries

- Capital markets
- Consumer web
- Ad tech
- Insurance



- Hard disks are so 20<sup>th</sup> century
- Data is an inflationary coefficient in the enterprise
- CPU is going massively multi-core
- Storage, especially memory, is dropping

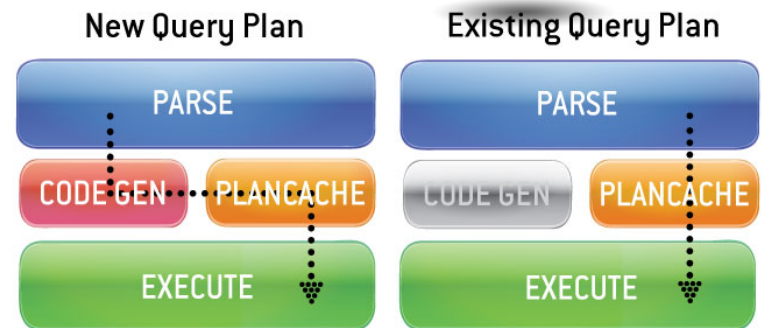




**TECHNOLOGY INNOVATIONS.**

---

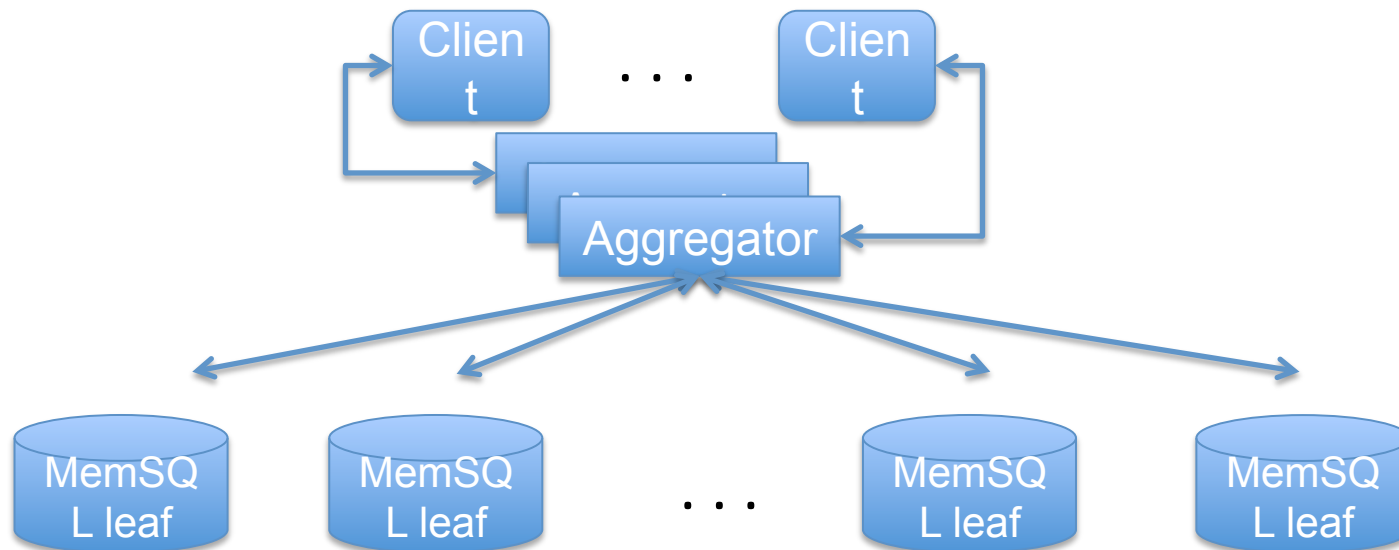
- Execute in parallel across the cluster on every core
- Squeeze each CPU cycle using SQL-to-machine code conversion for efficient execution
- Auto-parameterization keeps compilation to a minimum



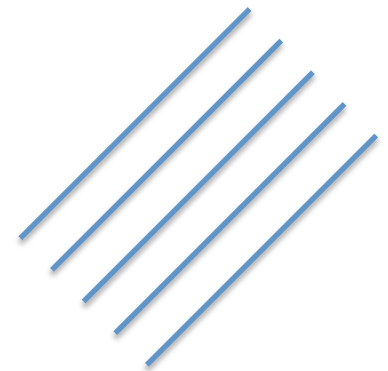
- Remove locks from the system
- No need for B-trees: use lock-free skip lists and hash tables for massive concurrency
- MVCC
- ACID compliant



- Shared-nothing architecture
- Distributed query optimizer
- Fault tolerant



- The goal: leverage hundreds/thousands of cores simultaneously.
- A distributed system to manage state is fundamental for successful infrastructure.
- Hadoop works, but it's
  - Not real-time
  - Hard to use for complex querying





# MODERN PROBLEMS, MODERN SOLUTIONS.

- Scale horizontally on **commodity** hardware
- RAM is the new disk; 1000+ core clusters are the norm
- It's good to be impatient: real-time analytics



**THANK YOU.**

---



**CONTACT**  
sales@memsql.com

**WEB**  
www.memsql.com

**425 2<sup>nd</sup> ST**  
San Francisco, CA 94107

**200 Park S Ave**  
New York, NY 10003