From Spark to Ignition:

Fueling Your Business on Real-Time Analytics

Eric Frenkiel, MemSQL CEO

June 29, 2015 • San Francisco, CA

What's in Store For This Presentation?

1. MemSQL:
   A real-time database for transactions and analytics
2. Spark Use Cases
3. Example: Geospatial Enhancements

MemSQL Story

The real-time database for transactions and analytics

MemSQL at a Glance

- Experienced leadership from Facebook, SQL Server, Oracle, Fusion-io
- In-Memory, distributed, relational database
- Solving the Enterprise Architecture Gap
- Horizontal scale-out with modern database innovation
- $50 million in funding

ACCEL PARTNERS khosla ventures IQT IN-Q-TEL

Data Collective First Round IA

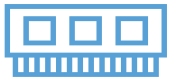
Four Ways Your DBMS is Holding You Back

- ETL (Extract, Transform, Load)
- Analytic Latency
- Synchronization
- Copies of data

Source: Gartner Hybrid/Transactional/Analytical Processing Will Foster Opportunities for Dramatic Business Innovation

The Real-Time Database for Transactions and Analytics

memsql



In-Memory

Distributed

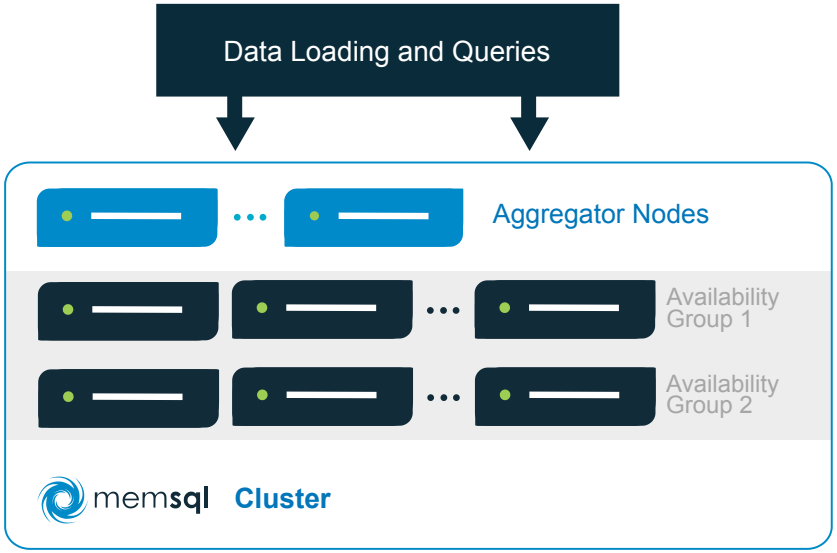
Relational

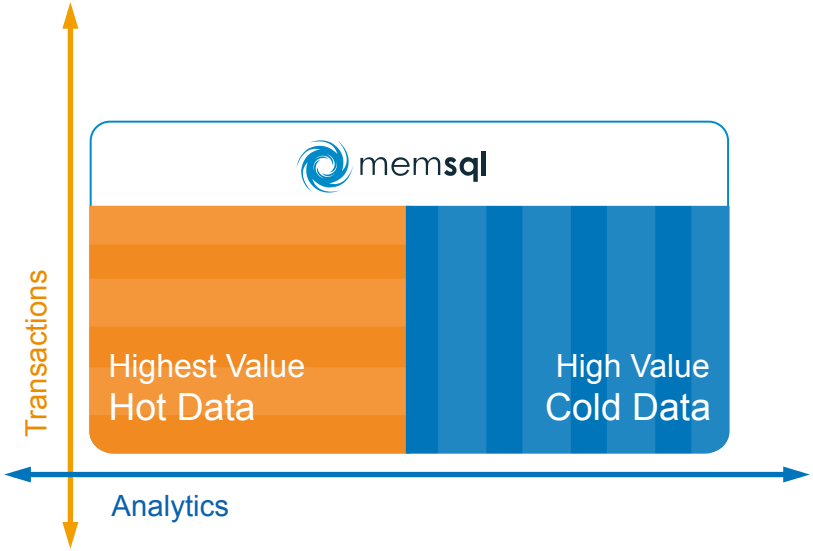
Software

Data Center

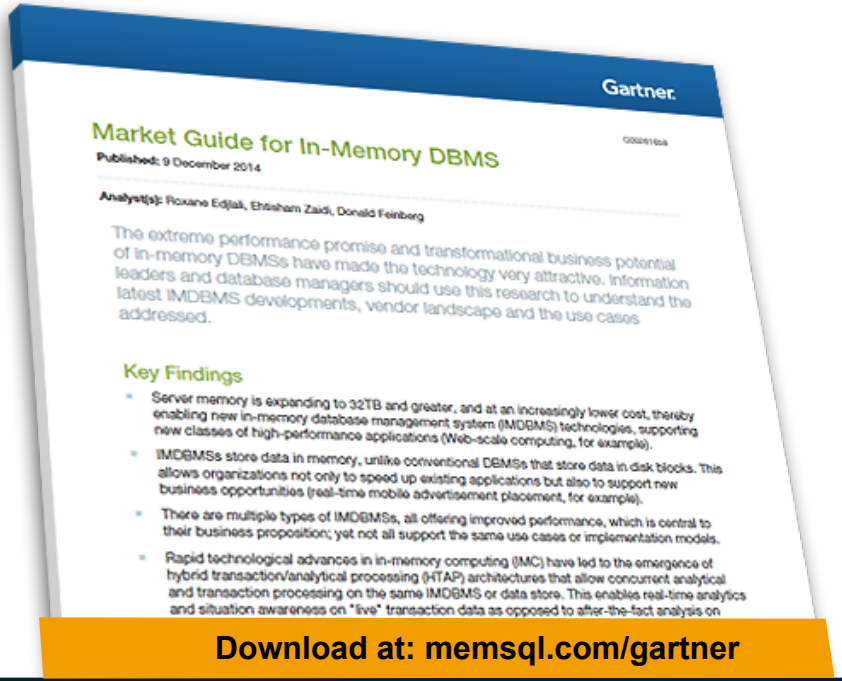
Cloud

The Real-Time Database for Transactions and Analytics





Gartner Identifies Emerging Category: HTAP (Hybrid Transactional/Analytical Processing)



Gartner

Market Guide for In-Memory DBMS

Published: 9 December 2014

Analyst(s): Roxane Edjlali, Ehtisham Zaidi, Donald Feinberg

The extreme performance promise and transformational business potential of in-memory DBMSs have made the technology very attractive. Information leaders and database managers should use this research to understand the latest IMDBMS developments, vendor landscape and the use cases addressed.

Key Findings

- Server memory is expanding to 32TB and greater, and at an increasingly lower cost, thereby enabling new in-memory database management system (IMDBMS) technologies, supporting new classes of high-performance applications (Web-scale computing, for example).
- IMDBMSs store data in memory, unlike conventional DBMSs that store data in disk blocks. This allows organizations not only to speed up existing applications but also to support new business opportunities (real-time mobile advertisement placement, for example).
- There are multiple types of IMDBMSs, all offering improved performance, which is central to their business proposition; yet not all support the same use cases or implementation models.
- Rapid technological advances in in-memory computing (IMC) have led to the emergence of hybrid transaction/analytical processing (HTAP) architectures that allow concurrent analytical and transaction processing on the same IMDBMS or data store. This enables real-time analytics and situation awareness on "live" transaction data as opposed to after-the-fact analysis on

Download at: memsql.com/gartner

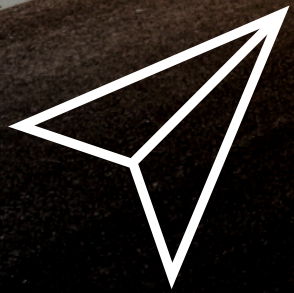
"HTAP will enable business leaders to perform…much more advanced and sophisticated real-time analysis of their business data than with traditional architectures."

Simple

- Standard SQL
- Transactions and analytics in one database
- Behind the firewall or on the cloud
- Flexible integrations (Hadoop, Spark, SQL)

Fast

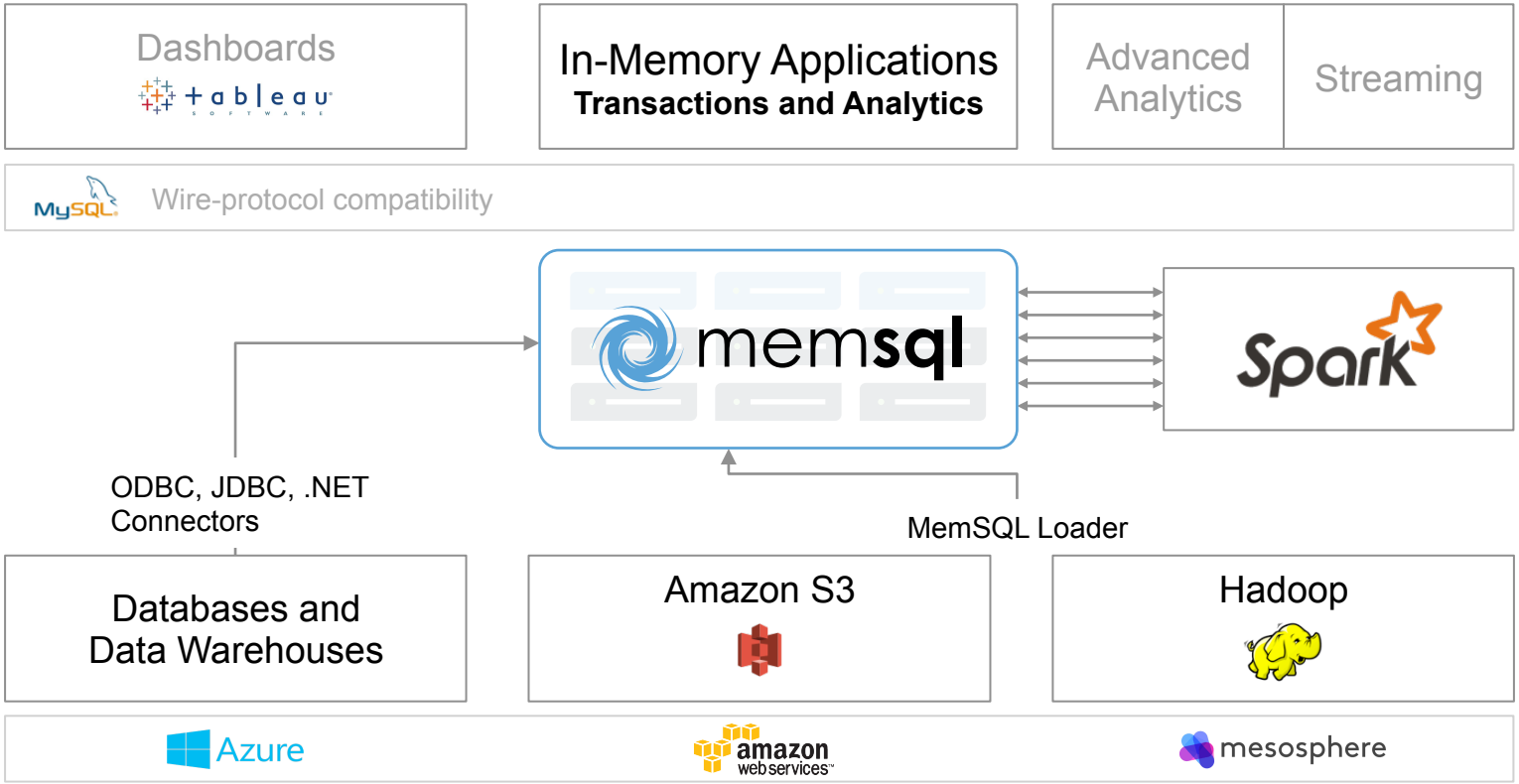
- Extremely low-latency queries
- Massive parallel transaction capacity
- Lock-free, shared-nothing architecture



Scalable

- Scales out on cloud and commodity hardware
- Deploys to thousands of machines
- True linear scaling

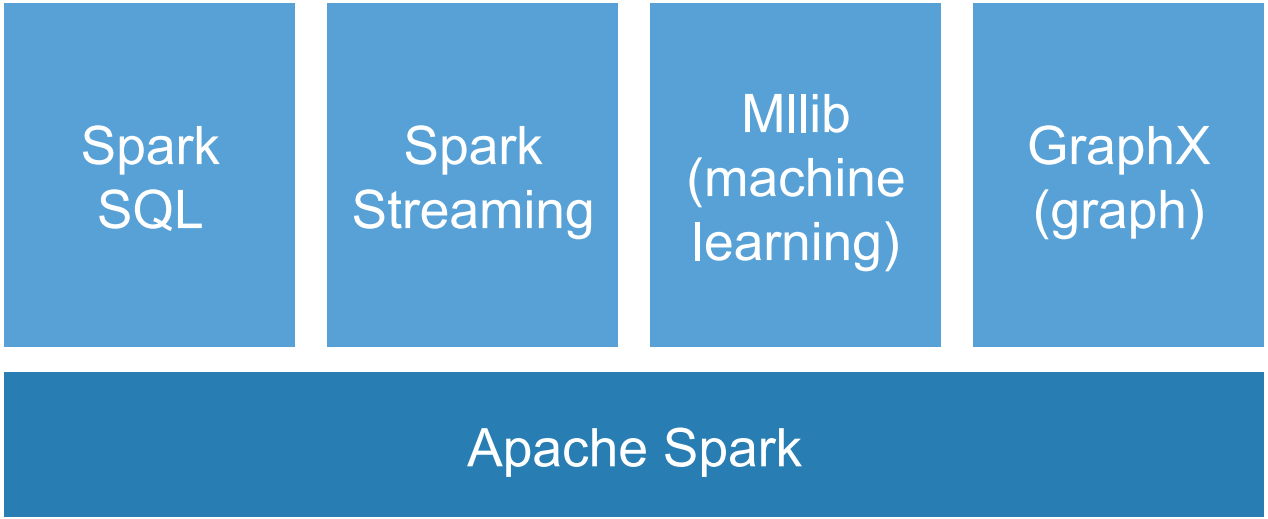
MemSQL Product Ecosystem



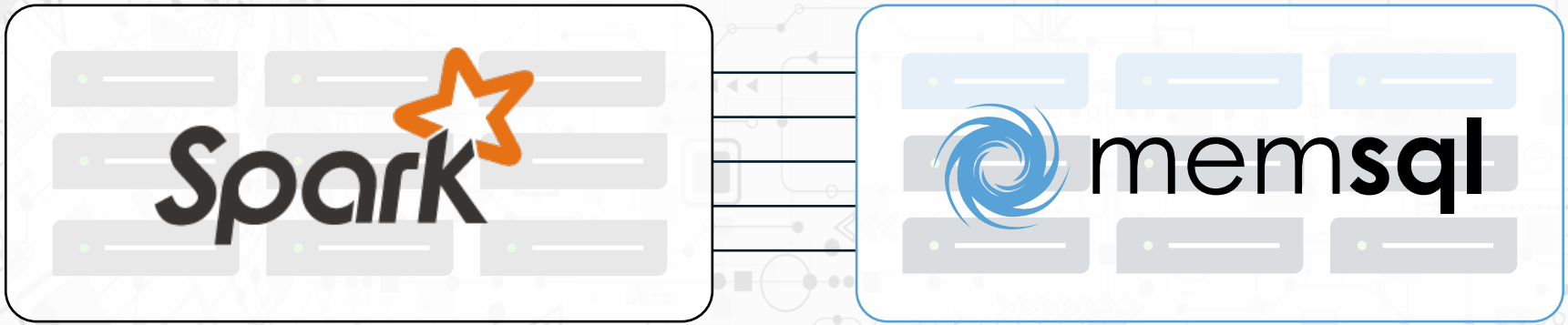
Spark Use Cases

Spark Data Processing Framework

Intuitive, concise, and expressive operations needed for analytics



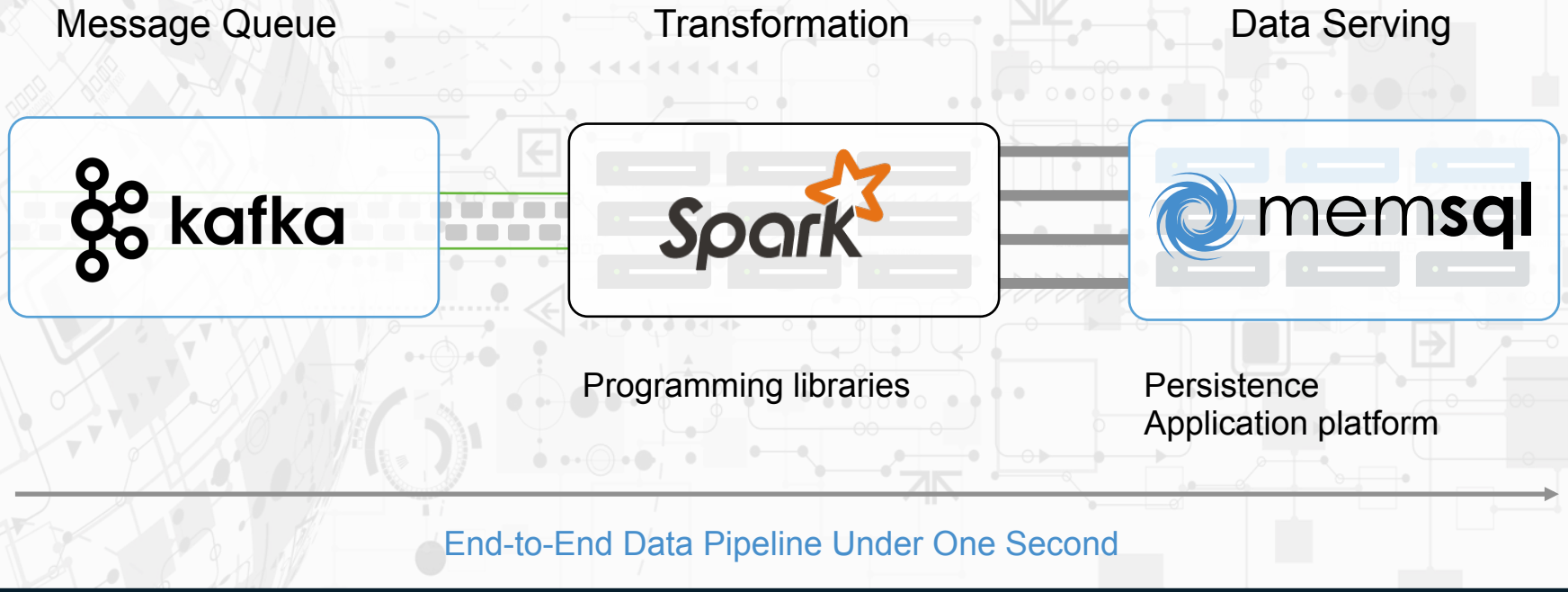
Understanding MemSQL and Spark



Cluster-wide Parallelization | Bi-Directional

Spark with MemSQL

MemSQL Spark Connector enables the real-time trinity

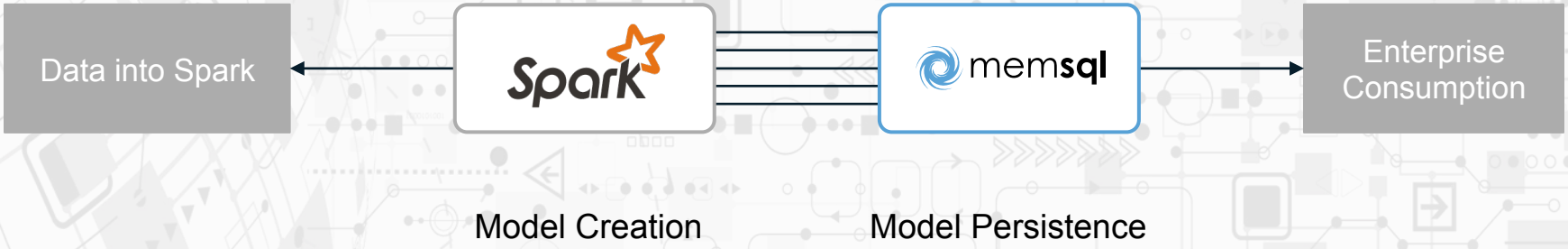


MemSQL and Spark Use Cases

- Operationalize models built in Spark
- Stream and event processing
- Live dashboards and automated reports
- Extend MemSQL analytics

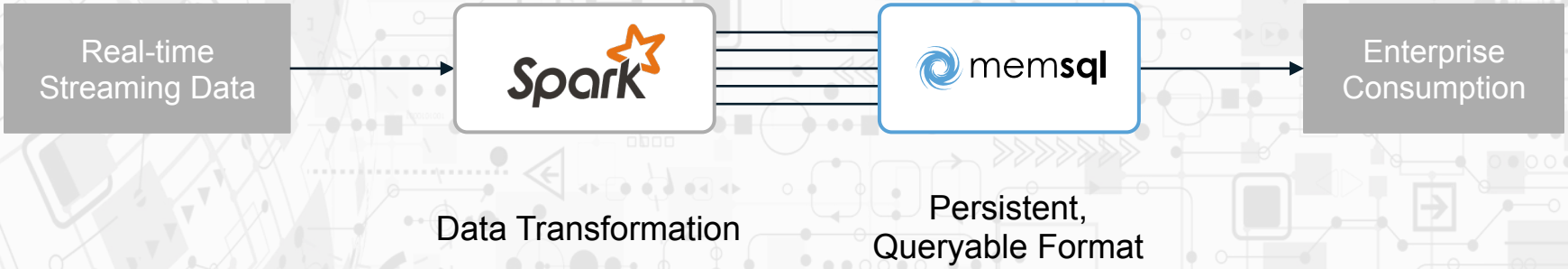
Operationalize Models Built in Spark

- Process in Spark, persist to MemSQL
- Go to production and iterate faster



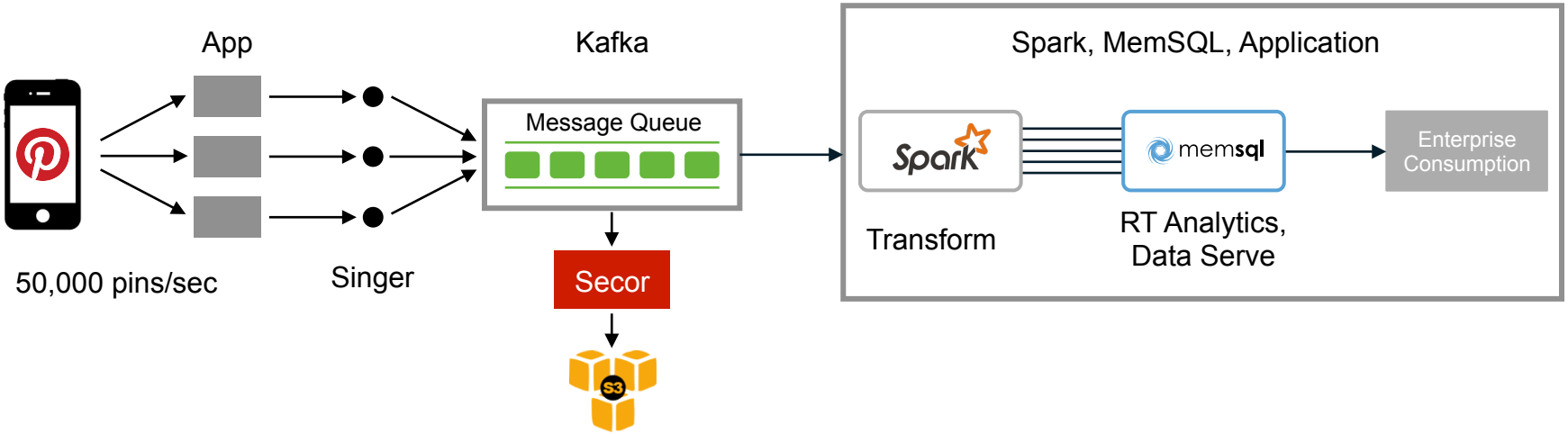
Stream and Event Processing

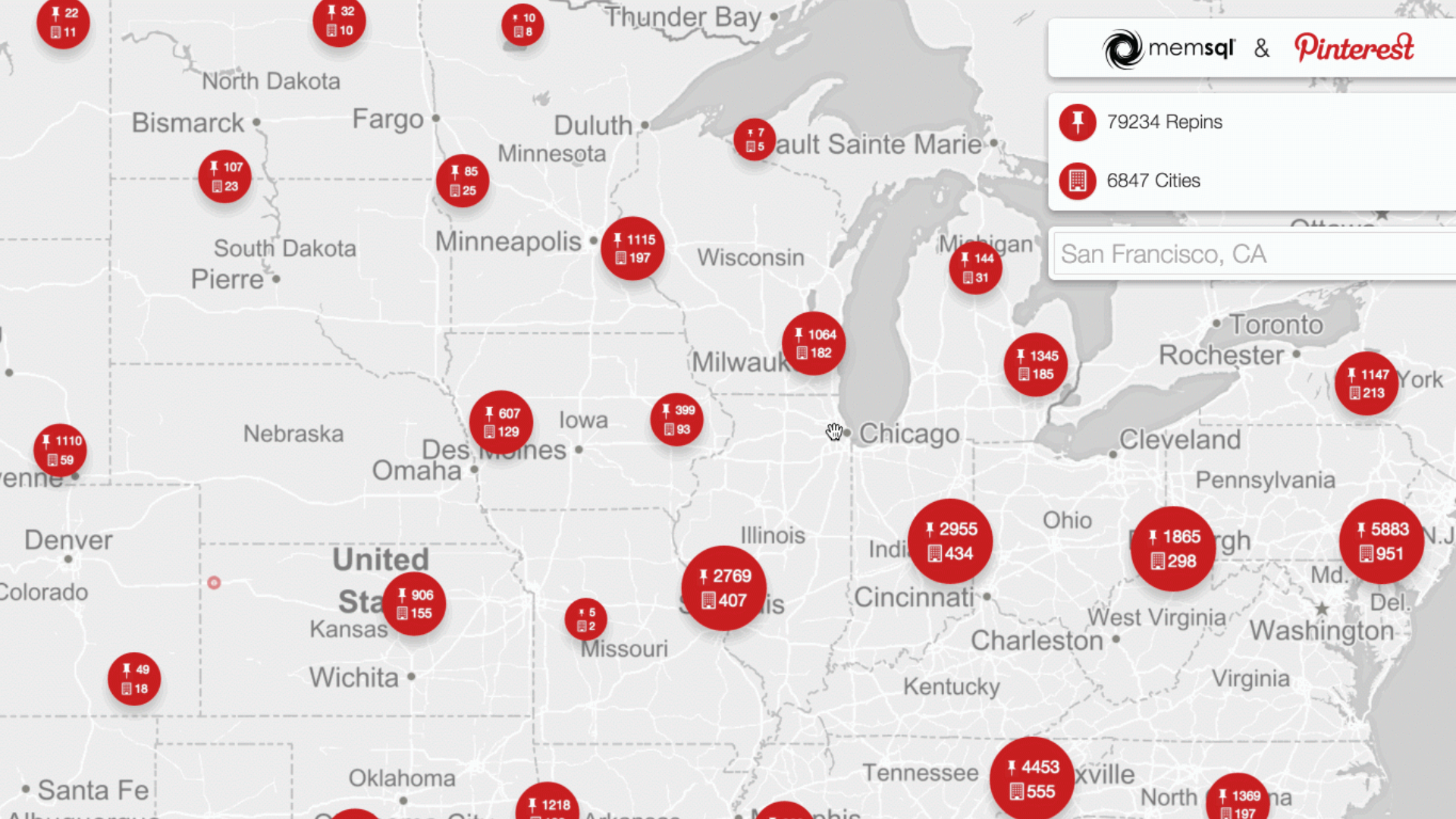
- Structure event data on the fly
- Pass to MemSQL for persistent, queryable format



Real-Time Analytics at Pinterest

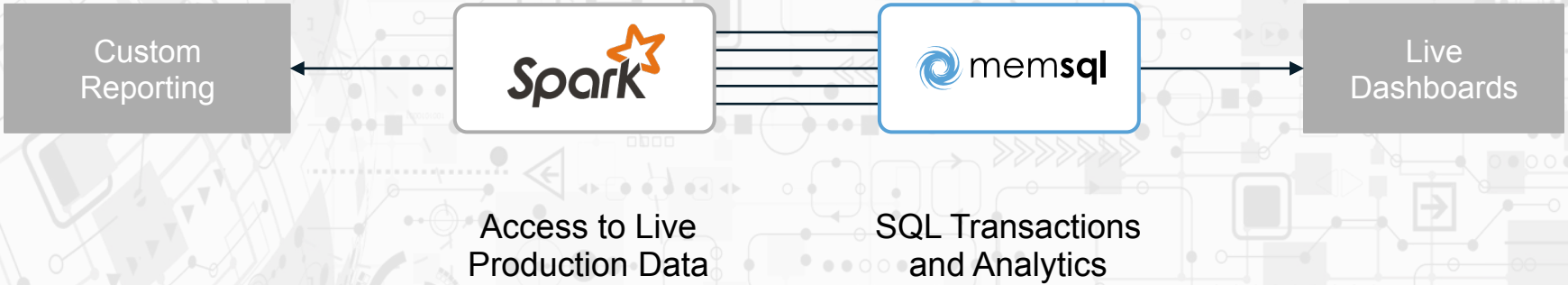
- Higher performance event logging
- Reliable log transport and storage
- Faster query execution on real-time data




memsql & Pinterest
79234 Repins
6847 Cities
San Francisco, CA
Thunder Bay
North Dakota
Bismarck
Fargo
Duluth
Minnesota
Minneapolis
South Dakota
Pierre
Wisconsin
Michigan
Milwaukee
Chicago
Toronto
Rochester
Cleveland
Pennsylvania
Ohio
Nebraska
Iowa
Des Moines
Omaha
Illinois
Indiana
Cincinnati
West Virginia
Charleston
Washington
Kentucky
Virginia
Tennessee
Denver
Colorado
Kansas
Missouri
Wichita
Oklahoma
Santa Fe
22 11
32 10
10 8
7 5
107 23
85 25
1115 197
144 31
1064 182
1345 185
1147 213
607 129
399 93
1110 59
2955 434
1865 298
5883 951
2769 407
906 155
5 2
49 18
4453 555
1218
1369 197

Live Dashboards and Automated Reports

- Serve live dashboards from MemSQL
- Run custom reports on live data with Spark



Extend MemSQL Analytics

- The freshest data for analysis in Spark
- Load from MemSQL to Spark and write results on return

MemCity

- Capturing energy consumption data from 1.4 million households
- 8 devices per household
- 186,000 events per minute
- AWS hardware costs at $2.35 per hour

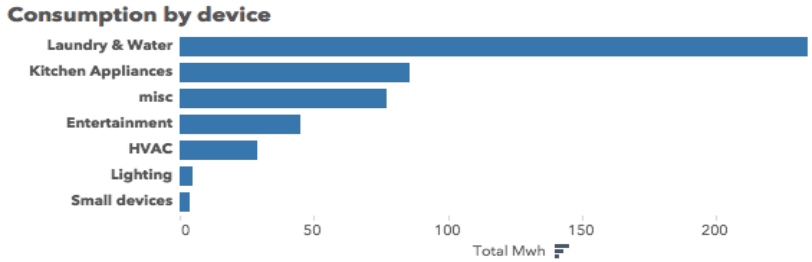
Kafka 322.0

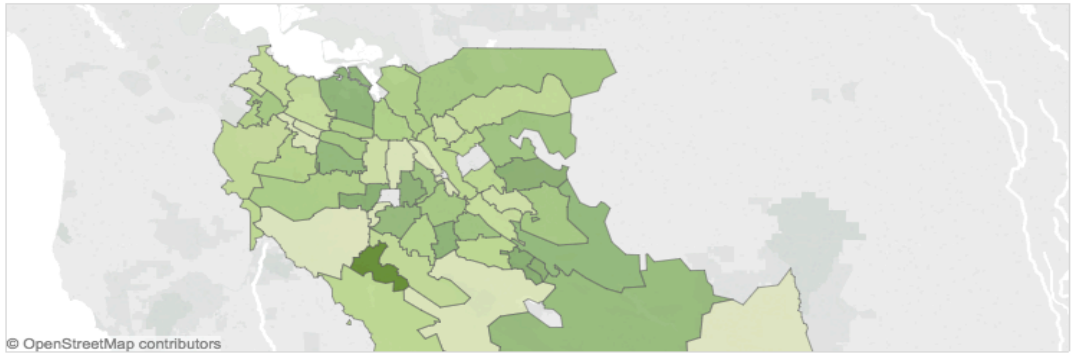
Spark 330.0

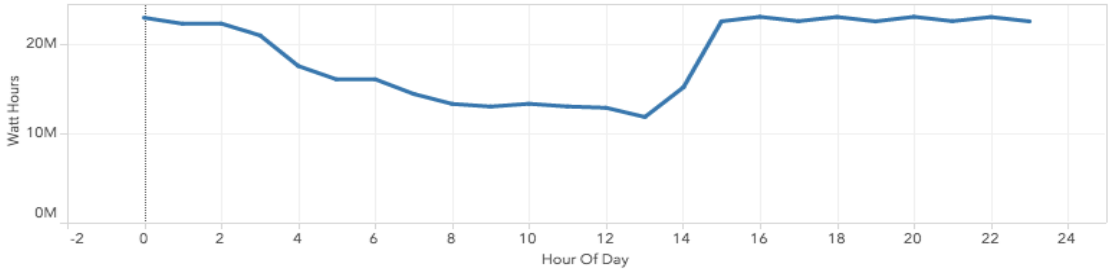
MemSQL 338.0

27,100,500,917

Megawatt-hours







141,041,035

Database records

kafka

Spark

memsql

+ableau SOFTWARE

Geospatial Enhancements

Geospatial Challenge

- Commercial applications *now geo-enabled*
- Location is everywhere
- Lots of insight possible
- Traditionally geo is processed separately
- Real need for integrated geospatial at scale

MemSQL Geospatial

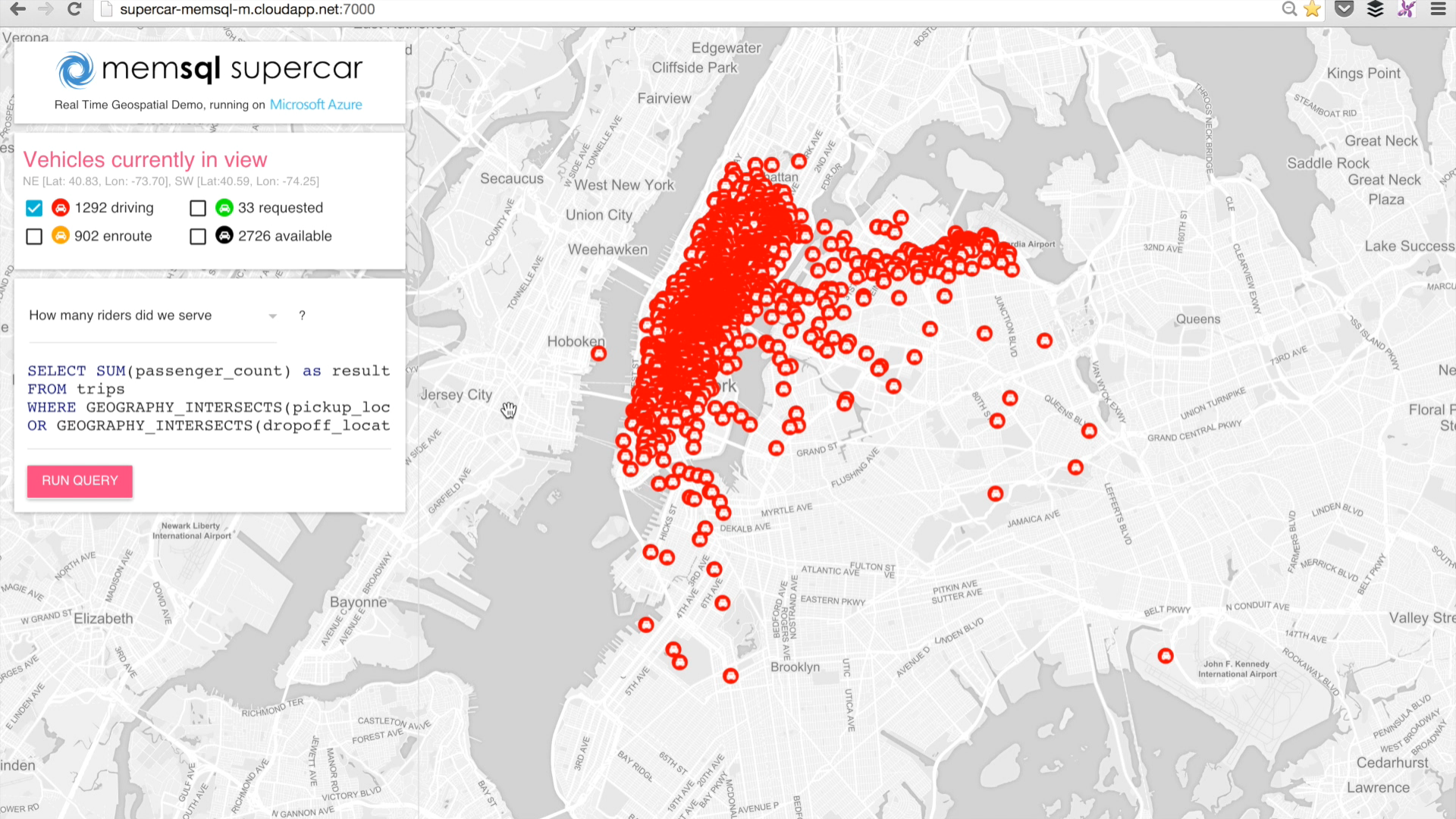
- Points, Lines, and Polygons
- Topological filters
- Measurement functions

MemSQL Geospatial

- BILLIONS of objects
- Sub-second latency
- Geo data is first-class citizen
- Geo + Simplicity + Speed + Scale

Real-Time Geospatial Location Intelligence

- Sample from 170 million taxi trips
- Real-time ingest
- Concurrent queries in fractions of a second
- Unlimited number of geographic views
- Simple queries while simultaneously ingesting data


supercar-memsql-m.cloudapp.net:7000
memsql supercar
Real Time Geospatial Demo, running on Microsoft Azure
Vehicles currently in view
NE [Lat: 40.83, Lon: -73.70], SW [Lat:40.59, Lon: -74.25]
1292 driving
33 requested
902 enroute
2726 available
How many riders did we serve ?
SELECT SUM(passenger_count) as result
FROM trips
WHERE GEOGRAPHY_INTERSECTS(pickup_loc
OR GEOGRAPHY_INTERSECTS(dropoff_locat
RUN QUERY
Edgewater
Cliffside Park
Fairview
Secaucus
West New York
Union City
Weehawken
Hoboken
Jersey City
Kings Point
Great Neck
Saddle Rock
Great Neck Plaza
Lake Success
Queens
Brooklyn
Bayonne
Elizabeth
Newark Liberty International Airport
John F. Kennedy International Airport
Cedarhurst
Lawrence

MemSQL 4 Community Edition

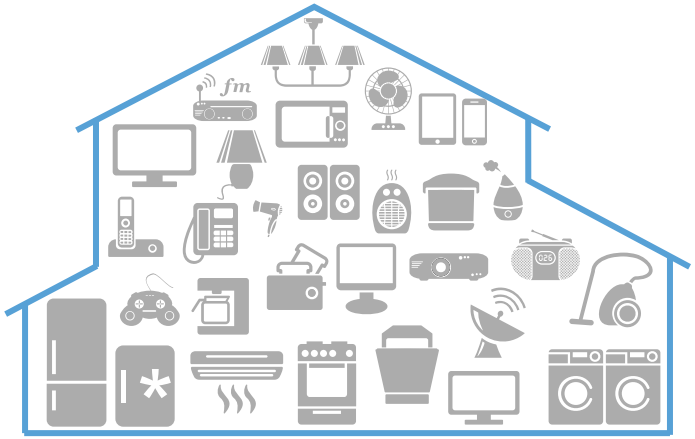
A database so scalable that everyone can use it.

UNLIMITED scale and capacity

Free FOREVER

Thank You!

Visit the MemSQL Booth **#4**



MemCity Showcase

Games

Giveaways