

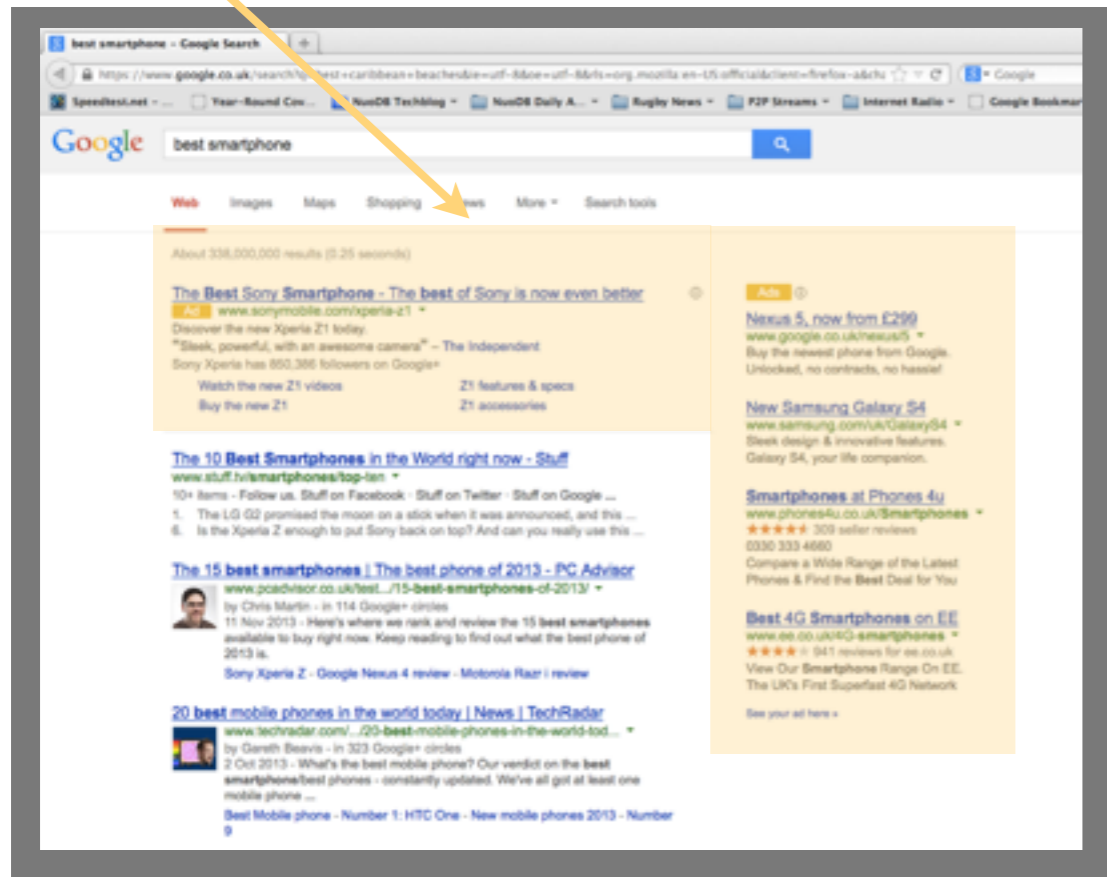


# The Future of Distributed Databases

Barry Morris, NuoDB Inc

# GOOGLE AdWords

## AdWords



### AdWords Background:

- ▶ GOOGLE's Primary Revenue Source
- ▶ Pay-per-click advertising
- ▶ Based on search string content
- ▶ \$42.5bn in 2012

### Very Demanding Application:

- ▶ Multiple Apps/Single Database
- ▶ Elastic Capacity On-demand
- ▶ Geo-Distributed
- ▶ Extreme Transactions, Analytics, Concurrency
- ▶ Continuous Availability



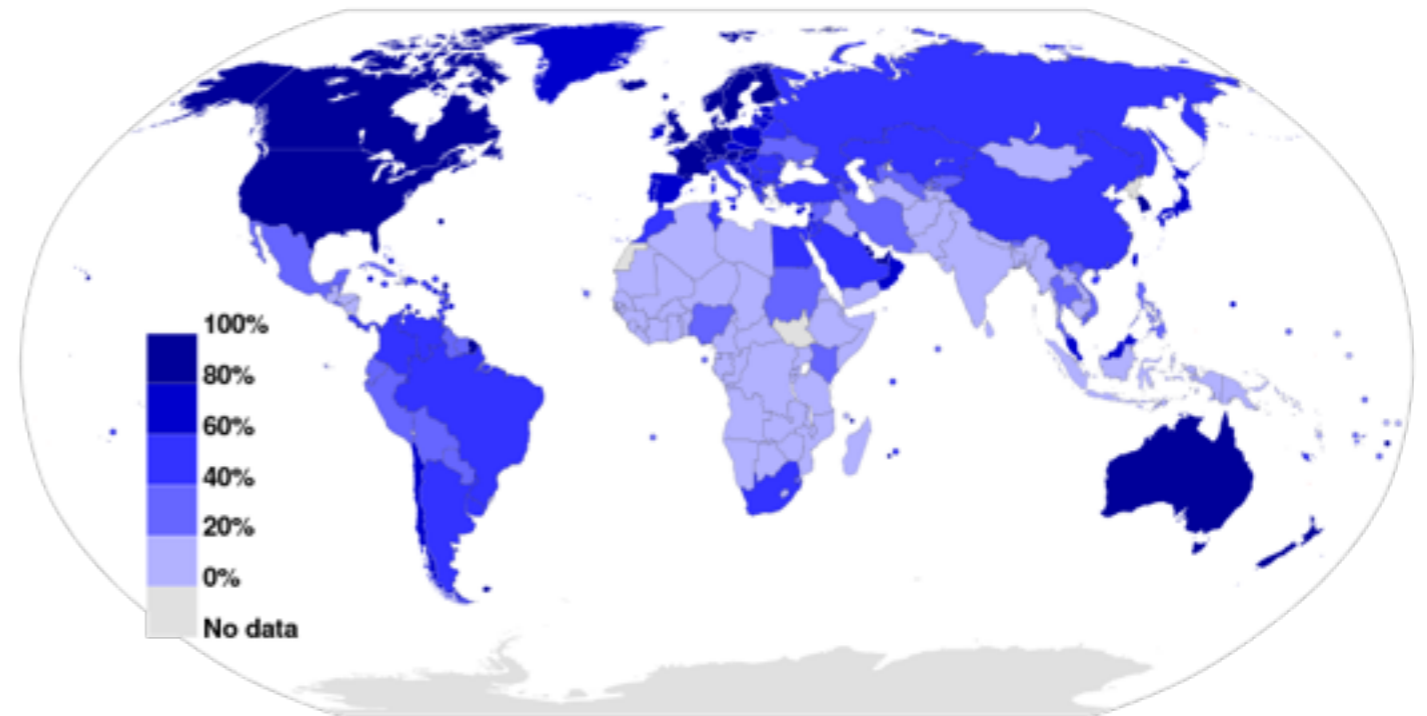
"When we sought a replacement for Google's MySQL data store for the AdWords product, [a KV Store] was simply not feasible: the complexity of dealing with a non-ACID data store in every part of our business logic would be too great, and there was simply no way our business could function without SQL queries. **Instead of going NoSQL, we built F1.**"

- GOOGLE F1 White Paper, 2013

# AdWords Epitomises NextGen Apps

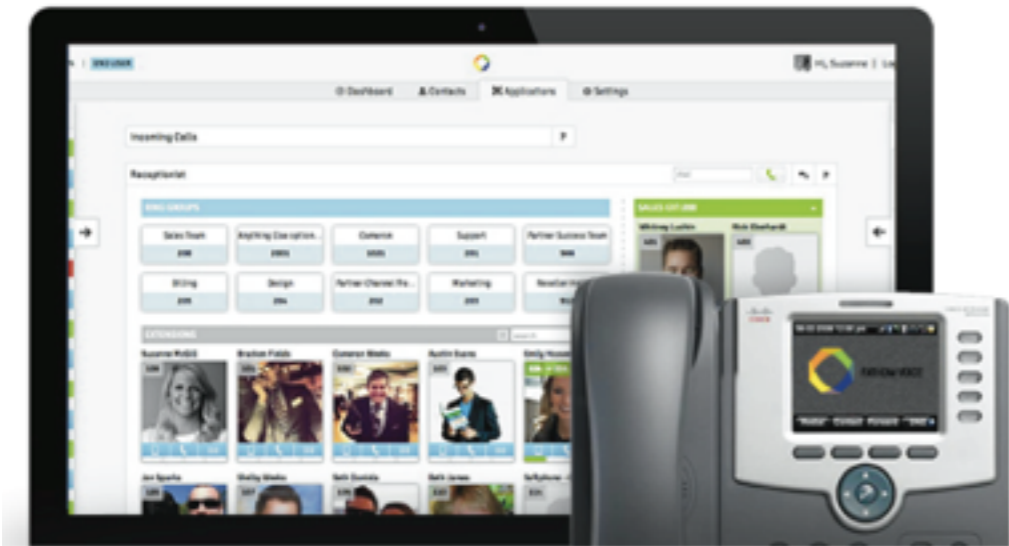
**Q:** How many other apps are targeted at a global audience, with requirements for elastic capacity on-demand, transactional and analytical workloads, geo-distributed requirements, and continuous availability?

**A:** Most apps are headed that way.



- ▶ 40% of the world's population on the internet today - Most developed countries it is over 80%
- ▶ \$75bn devices on the internet by 2020: Every car, watch, refrigerator, TV set and toothbrush
- ▶ Every WebApp, Mobile App and IoT app will be used in a geo-distributed fashion, with transactions, analytics and continuous availability.
- ▶ In other words **AdWords-like applications will be everywhere**

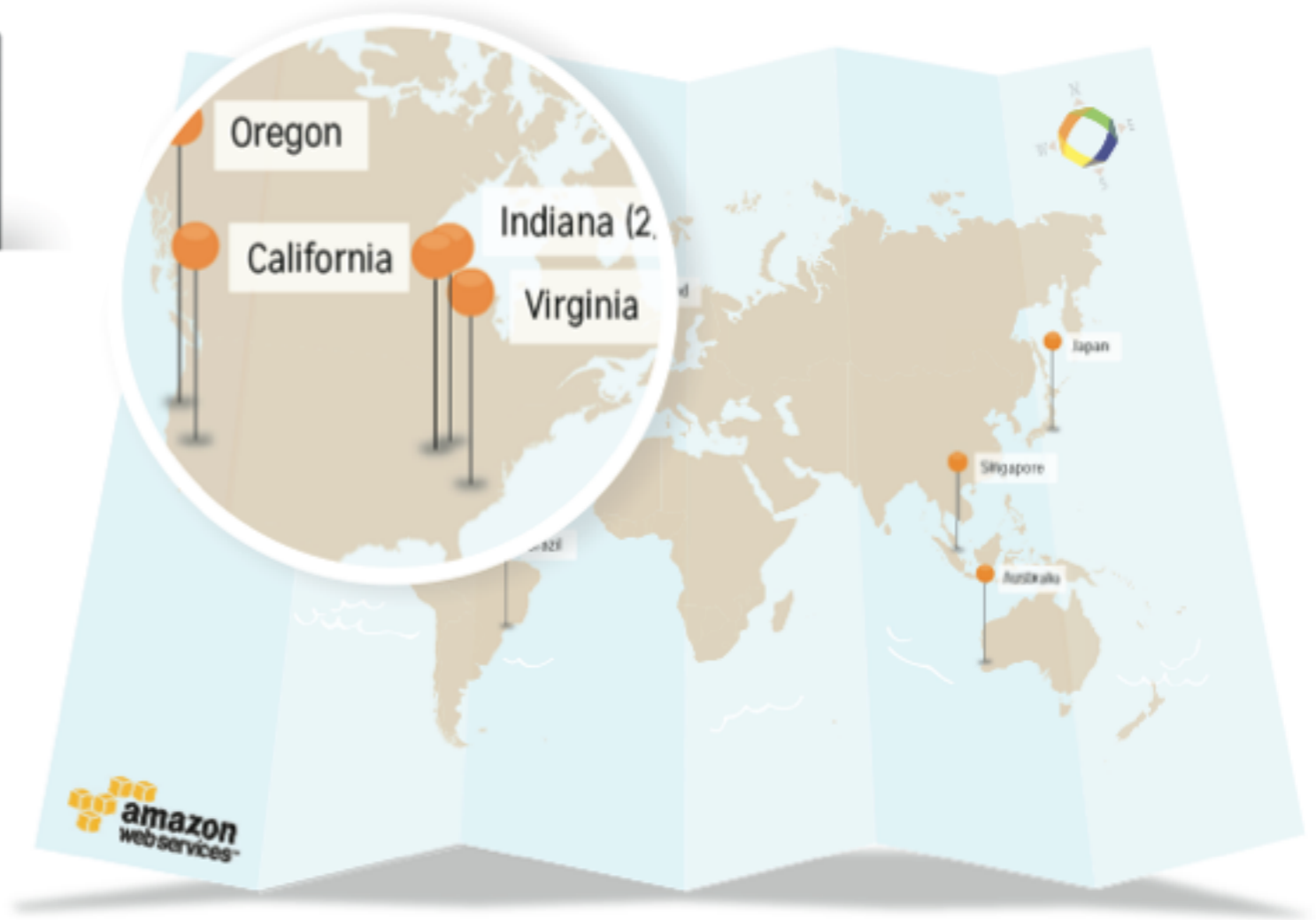
# Customer: Fathom Voice



The Receptionist application inside the industry leading web portal. Go.FathomVoice

## **Fathom Voice**

Fathom Voice is a communications software company focused on building industry-leading technology. Their flagship product is a cloud-based phone system that offers hundreds of advanced calling features and an industry-leading web portal, Go.FathomVoice. With Fathom Voice, the capabilities of the PBX phone are enhanced and users are able to communicate better, faster and easier.



# Fathom Voice

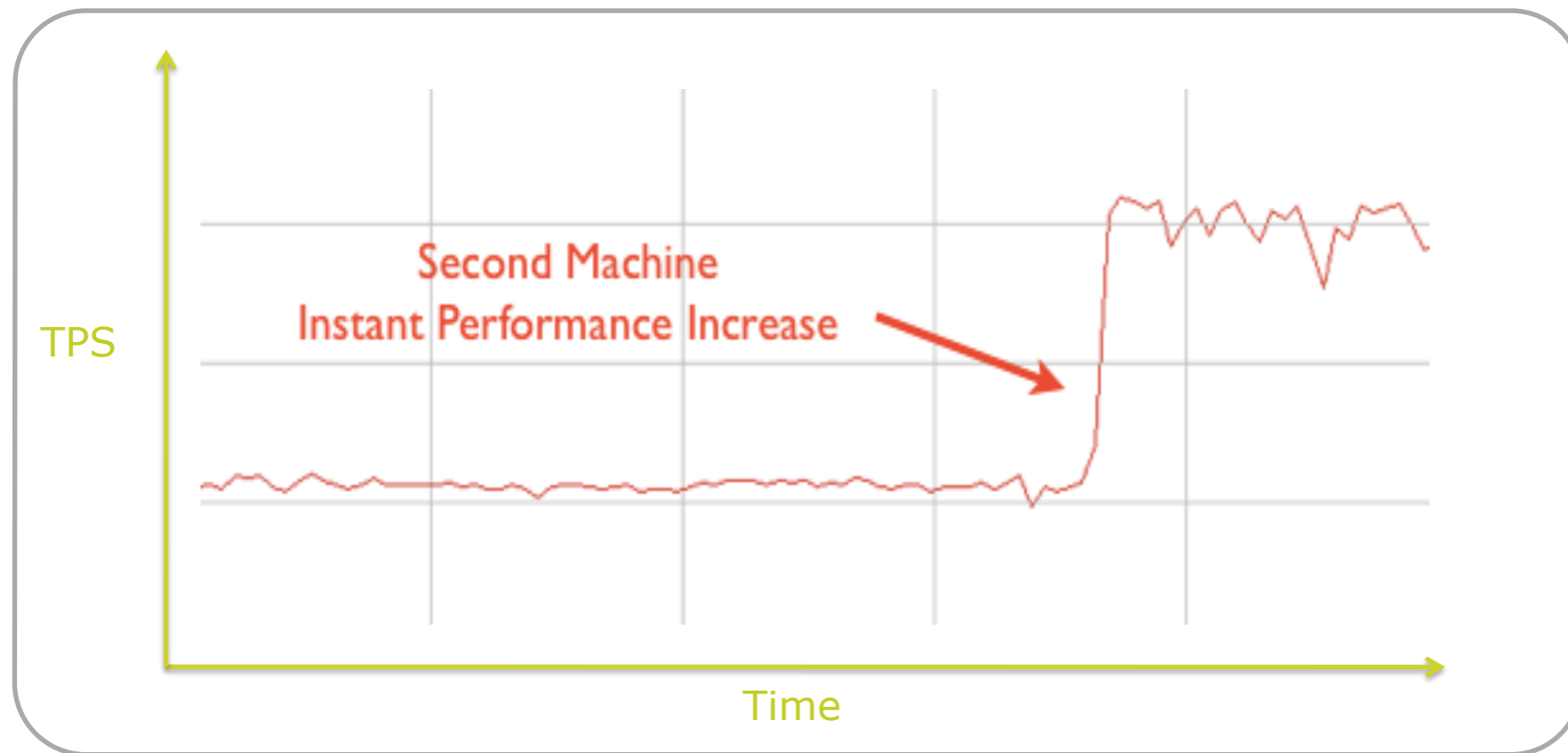


“We needed a single, logical database that could be shared across multiple servers in different geographies; updated in real-time; and automatically scaled out during peak demand to handle increased traffic, then back in during off-peak hours”

“We were told to stop trying to change the way data management works and conform our service to the existing database solutions. But, that’s not who we are.”

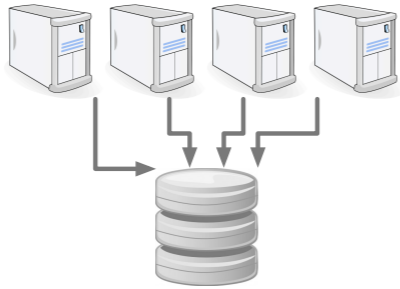
- Cameron Weeks, CEO Fathom Voice

# Distributed Transactions

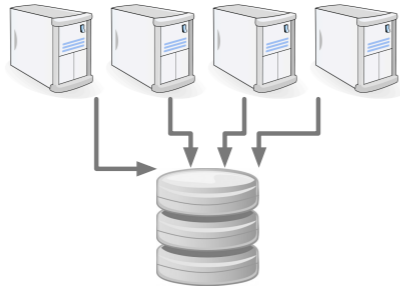
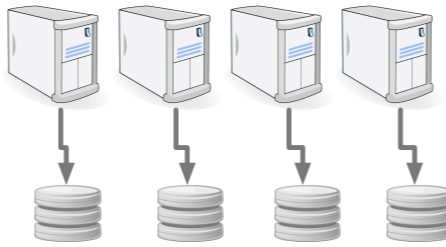


- ▶ Dynamically add/delete machines to manage capacity
- ▶ No sharding, no replication, no memcached
- ▶ Resilience to failure
- ▶ Single logical database
- ▶ Geo-distributed Operation

# Distributed Transactions: Designs

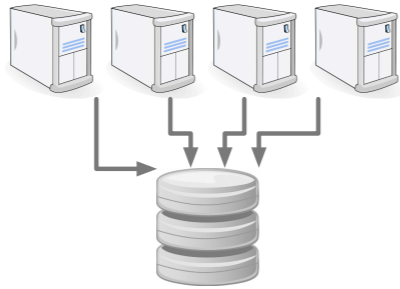
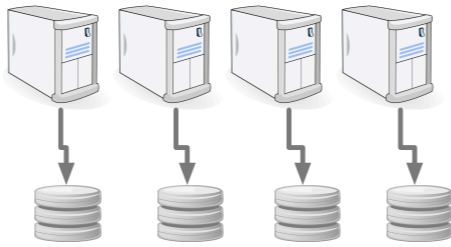
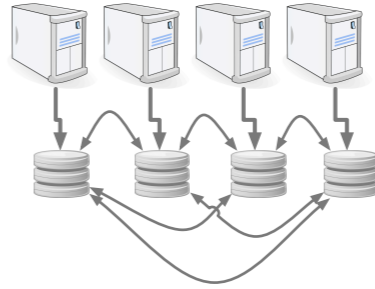

Approach	Shared Disk
Key Idea	Sharing a file system.
Topology	
Examples	ORACLE RAC, Tandem Nonstop SQL, MS Sql Server Cluster, ScaleDB

# Distributed Transactions: Designs

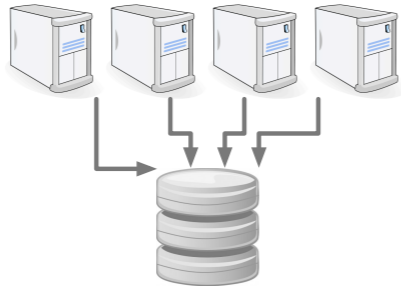
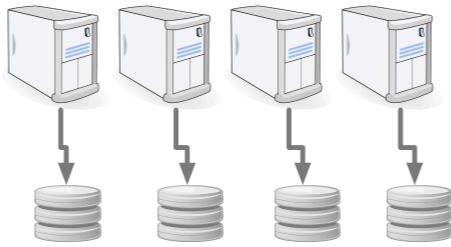
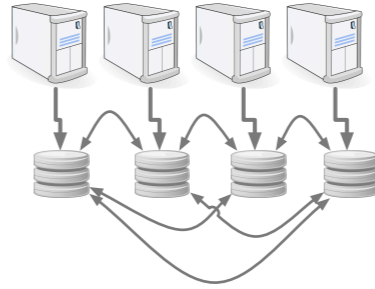
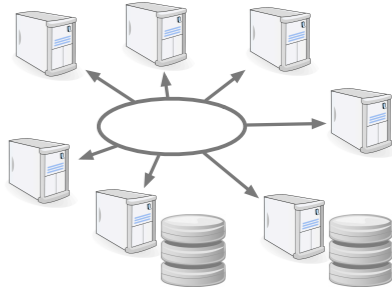


Approach	Shared Disk	Shared-Nothing/ Sharded
Key Idea	Sharing a file system.	Independent databases for disjoint subsets of the data.
Topology		
Examples	ORACLE RAC, Tandem Nonstop SQL, MS Sql Server Cluster, ScaleDB	<p>Clustrix, VoltDB, MemSQL, Xkoto, ScaleBase, MongoDB, and most NoSQL/ NewSQL solutions.</p> <p>Note: Most major web properties include custom sharded MySQL or sharded PostgreSQL, including Facebook, GOOGLE, Wikipedia, Amazon, Flickr, Box,net, Heroku, ...</p>



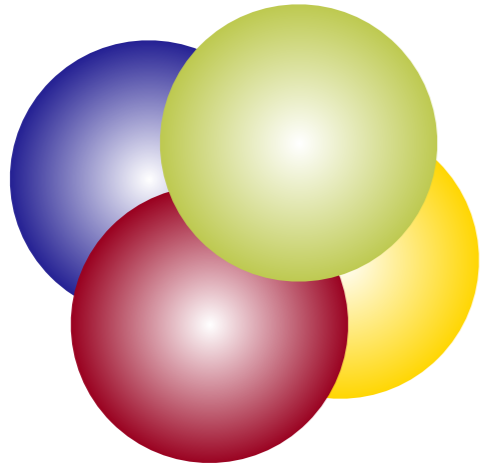
# Distributed Transactions: Designs

Approach	Shared Disk	Shared-Nothing/ Sharded	Synchronous Replication
Key Idea	Sharing a file system.	Independent databases for disjoint subsets of the data.	Committing data transactionally to multiple locations before returning.
Topology			
Examples	ORACLE RAC, Tandem Nonstop SQL, MS Sql Server Cluster, ScaleDB	Clustrix, VoltDB, MemSQL, Xkoto, ScaleBase, MongoDB, and most NoSQL/ NewSQL solutions.  Note: Most major web properties include custom sharded MySQL or sharded PostgreSQL, including Facebook, GOOGLE, Wikipedia, Amazon, Flickr, Box.net, Heroku, ...	

# Distributed Transactions: Designs

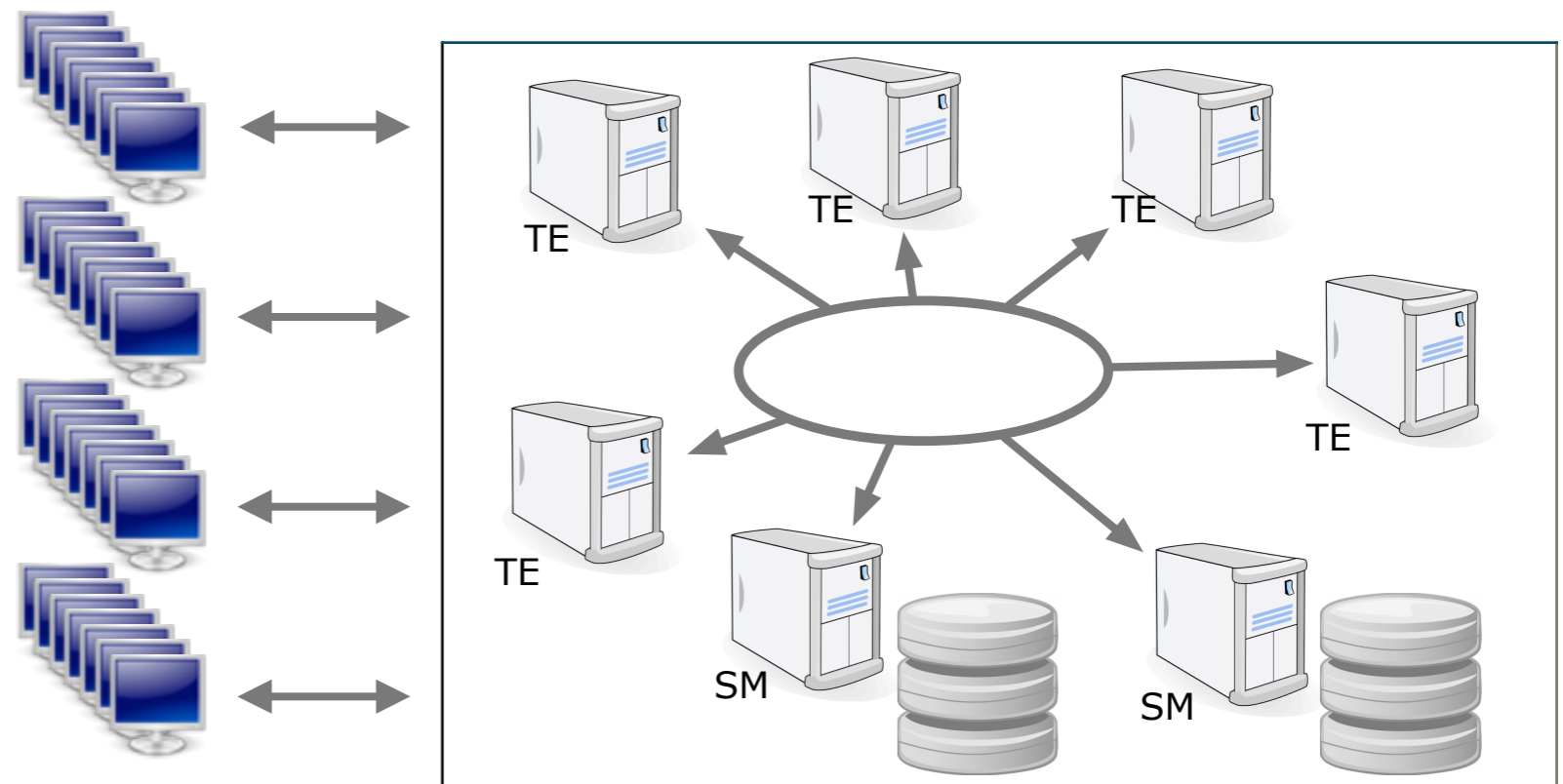
Approach	Shared Disk	Shared-Nothing/ Sharded	Synchronous Replication	Durable Distributed Cache
Key Idea	Sharing a file system.	Independent databases for disjoint subsets of the data.	Committing data transactionally to multiple locations before returning.	Replicating Data in memory on-demand.
Topology				
Examples	ORACLE RAC, Tandem Nonstop SQL, MS Sql Server Cluster, ScaleDB	Clustrix, VoltDB, MemSQL, Xkoto, ScaleBase, MongoDB, and most NoSQL/ NewSQL solutions.  Note: Most major web properties include custom sharded MySQL or sharded PostgreSQL, including Facebook, GOOGLE, Wikipedia, Amazon, Flickr, Box.net, Heroku, ...		

# Durable Distributed Cache



- ▶ All state is maintained as in-memory smart-objects called ATOMS
- ▶ ATOMS are loaded on demand and ejected when convenient (a la distributed cache)
- ▶ ATOMS are autonomous and communicate as peers
- ▶ ATOMS can serialize themselves to permanent backing stores
- ▶ ATOMS can maintain any number of replicas of themselves

- ▶ Add as many TEs/SMs as you like
- ▶ Distributed I/O, eg Hadoop HDFS
- ▶ No single point of failure
- ▶ Unlimited databases per Domain
- ▶ Single Console Management



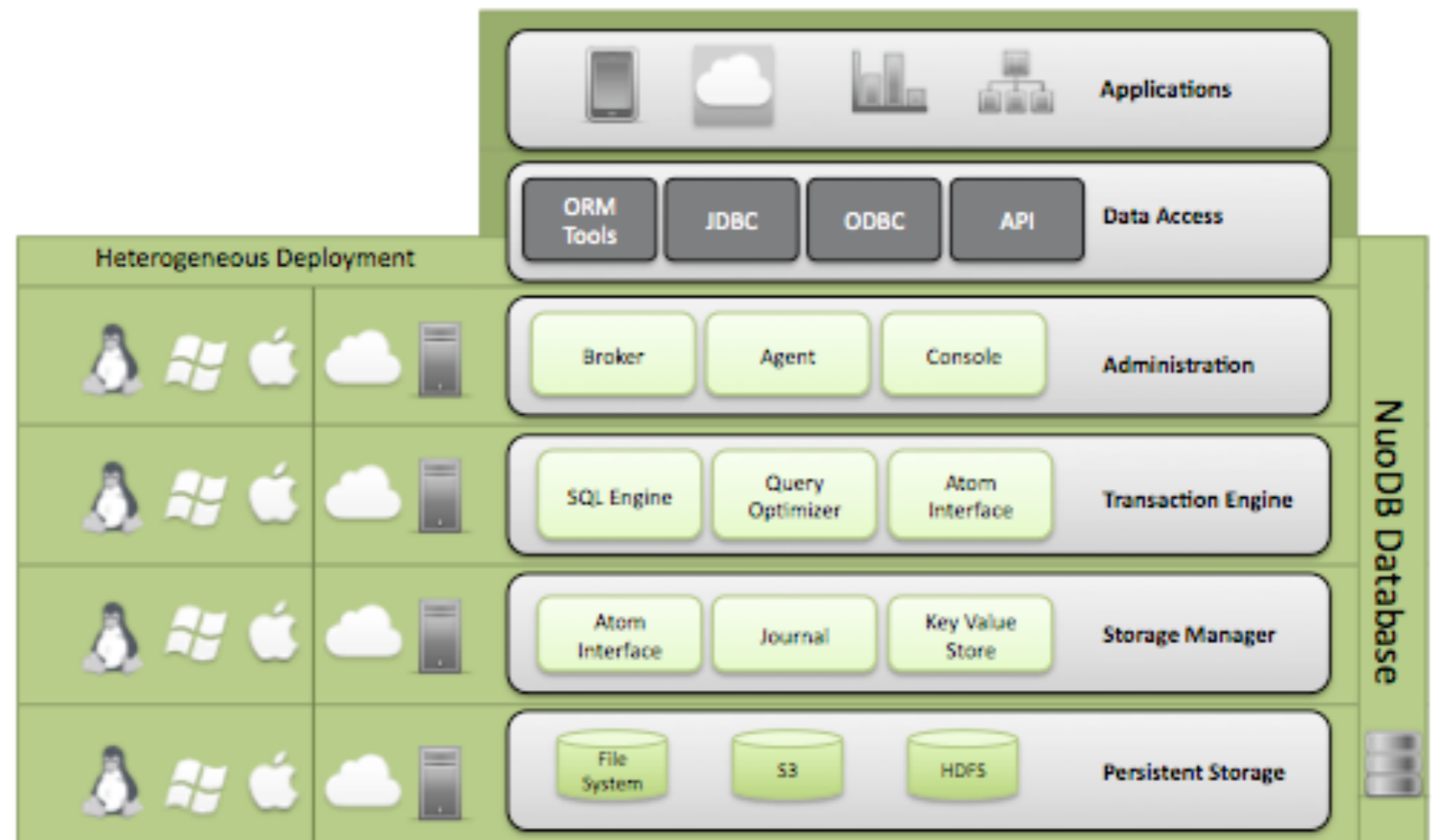
Applications

NuoDB Security/Admin Domain  
(TE = Transaction Engine, SM = Storage Manager)

# It's Not about SQL/NoSQL

Pluggable SQL/NoSQL Layer

Distributed Transaction Layer



The problem of distributed transactions is orthogonal to the choice of data model, language or access methods.

# Elastic Scalability

## Twitter:

- ▶ Over 140 million active users
- ▶ 4629 tweets per second (25,000 at peak)
- ▶ Three million new rows created per day
- ▶ 400 million tweets per day, replicated four times

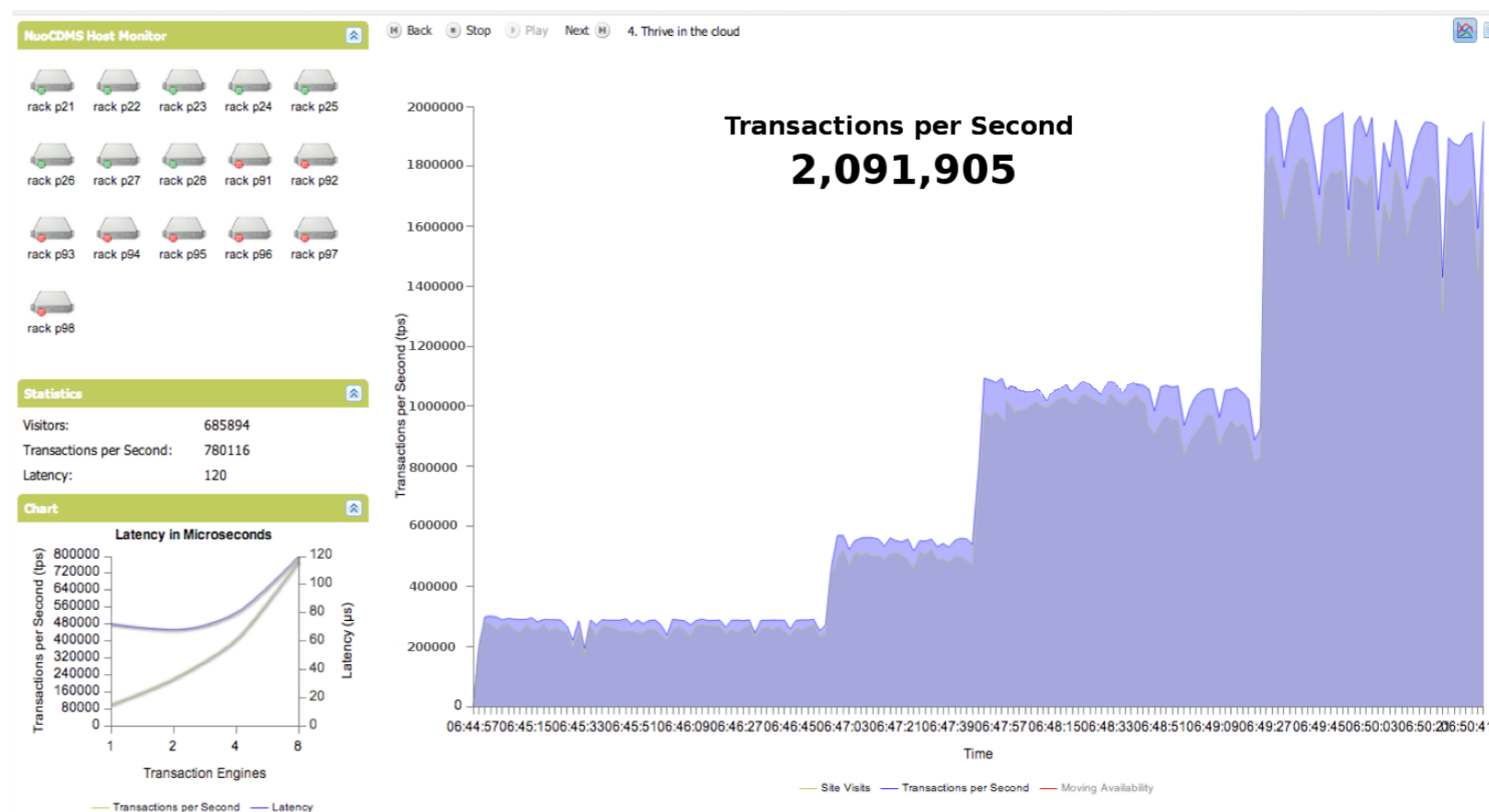
## Paypal:

- ▶ Over 100 million active users.
- ▶ 256-byte reads in under 10 milliseconds.
- ▶ Global replication of writes in under 350 milliseconds for a single 32-bit integer.
- ▶ Runs on Amazon Web Services in US, Japan and European data centres.

## Facebook:

- ▶ Over 950 million active users
- ▶ Rows read per second: 450 million (at peak)
- ▶ Queries per second: 13 million (at peak)
- ▶ Query response times: 4ms reads, 5ms writes
- ▶ Rows changed per second: 3.5 million (at peak)

(All 2011/2012 numbers)



- ▶ NuoDB scales to over 100 server machines
- ▶ Scalability is instant and elastic
- ▶ Scales-out and scales-in
- ▶ TPS numbers exceed 10m TPS on \$100k of hardware
- ▶ Also scales on AWS, GCE etc. Public demo of 32 nodes with GOOGLE
- ▶ Now showing linear scalability on TPC-C type workloads (DBT-2)
- ▶ Scalability demonstrated with heavier duty customer applications (eg Axway, Dassault Systèmes)



# Continuous Availability

 **Oracle database crashes JPMorgan Chase web site**  
Posted by Jeffrey Hebert on September 20, 2010 at 3:01pm  
[View Blog](#)

Last week, over 16 million customers of America's second largest bank were unable to access their accounts or process online payments for a good part of the week. Angry customers, hit by late charges for scheduled payments.

**RBS takes £125m hit over IT outage**  
by Dan Worth 03 Aug 2012  
More from this author 1 Comment

## Virgin Blue settles over check-in systems outage

By Liz Tay on Apr 5, 2011 5:11 PM  
Filed under: [Outages](#)

## GitHub Says Database Issues Caused This Week's Outage and Performance Problems

  
SITE STATUS

### PLANNED DATABASE MAINTENANCE – SAT 6/22/2013

On Saturday, June 22, 2013 from 10:00AM to 2:00PM PST, Yammer will be applying a major database software patch which may require approximately 4 hours of downtime. All precautions have been taken to ensure downtime will be as minimal as possible and no customer data will be lost.

As always, we appreciate your patience!



- ▶ Self-healing
- ▶ No single point of failure
- ▶ Fully distributed control
- ▶ Arbitrarily redundant:
  - ▶ Data
  - ▶ Control
  - ▶ Admin
- ▶ Online backup
- ▶ Online schema evolution
- ▶ Rolling upgrades

# Geo-Distribution



- ▶ Active/Active
- ▶ ACID Semantics
- ▶ Transactional Consistency
- ▶ N-Way Redundant
- ▶ Local User Latency
- ▶ Asynch WAN Comms

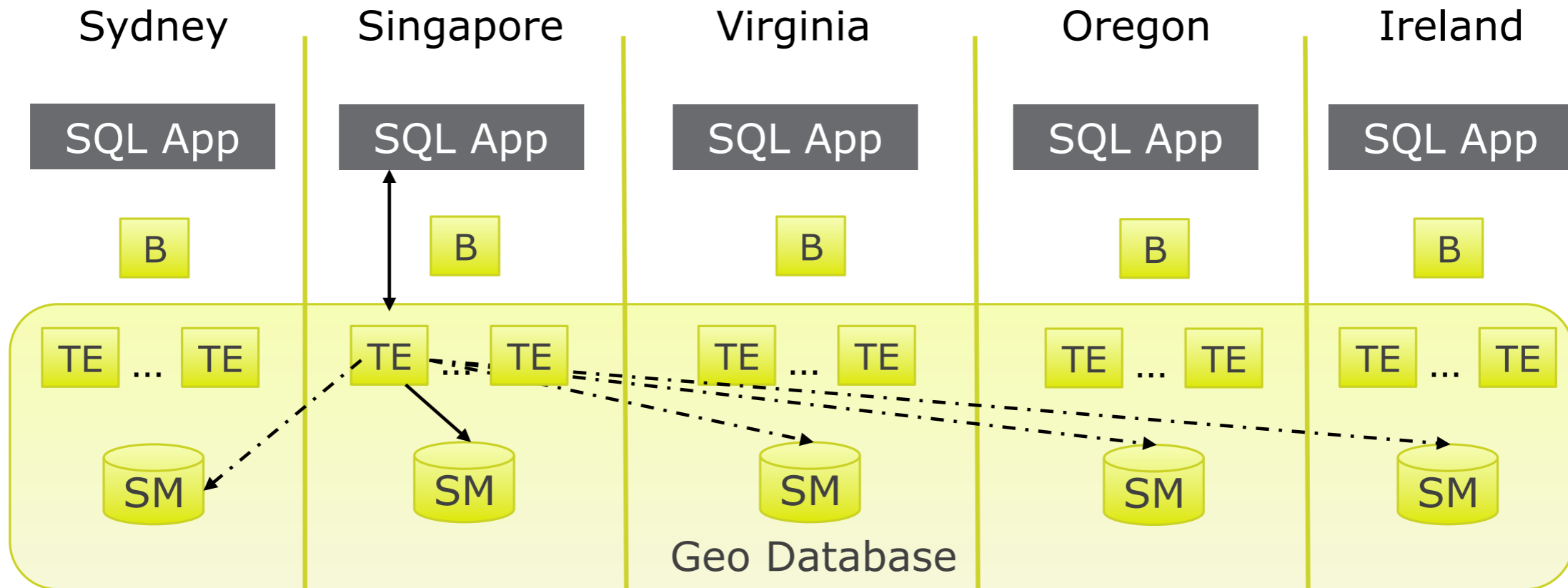
# Customer: Fathom Voice

“NuoDB is an entirely new database architecture, which replicates data in real time for an unlimited number of nodes. That is the most important piece for us. We can have all of our data everywhere and it can be updated in real-time.”

- *Cameron Weeks, CEO Fathom Voice*

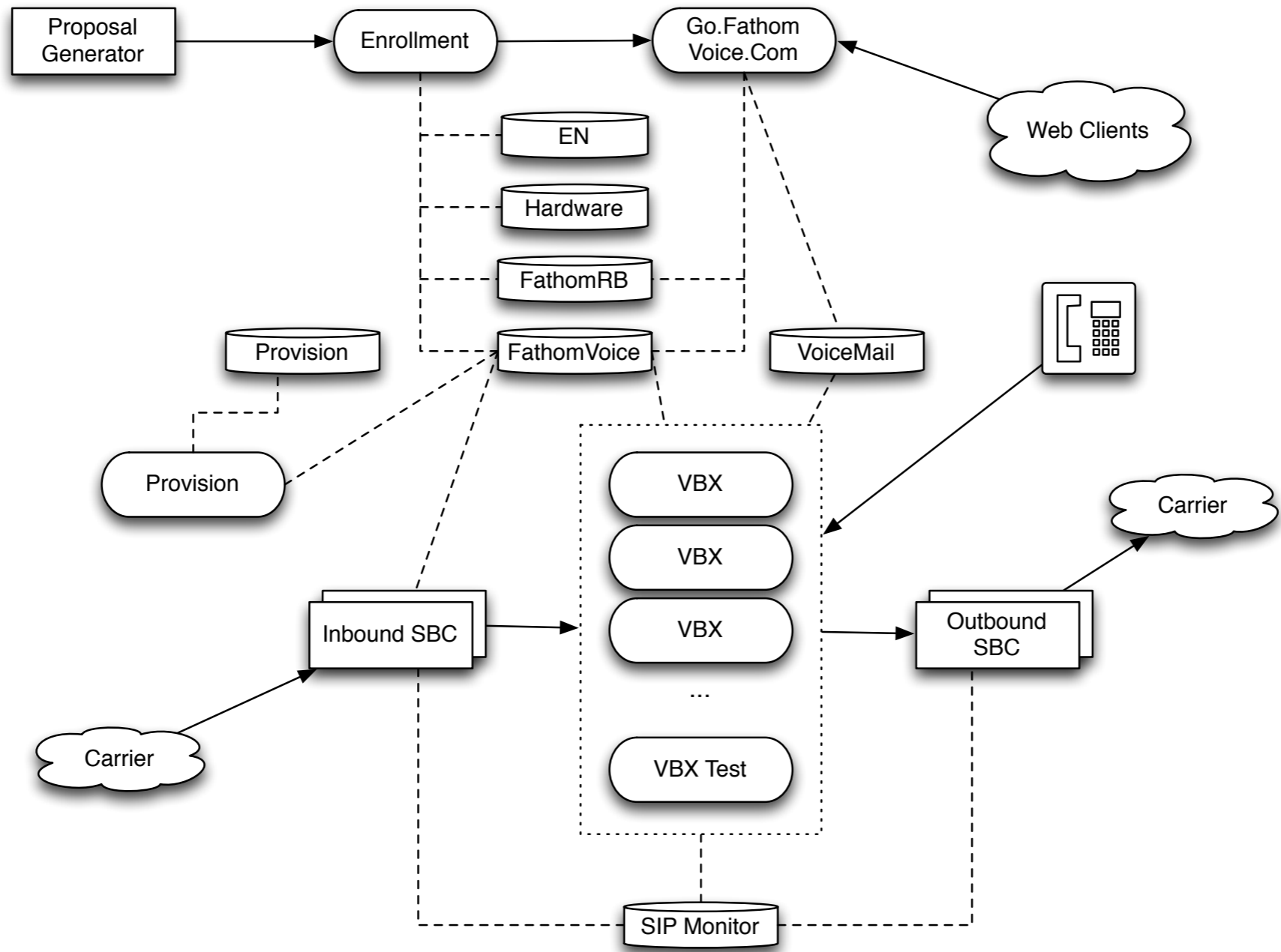


# Customer: Fathom Voice

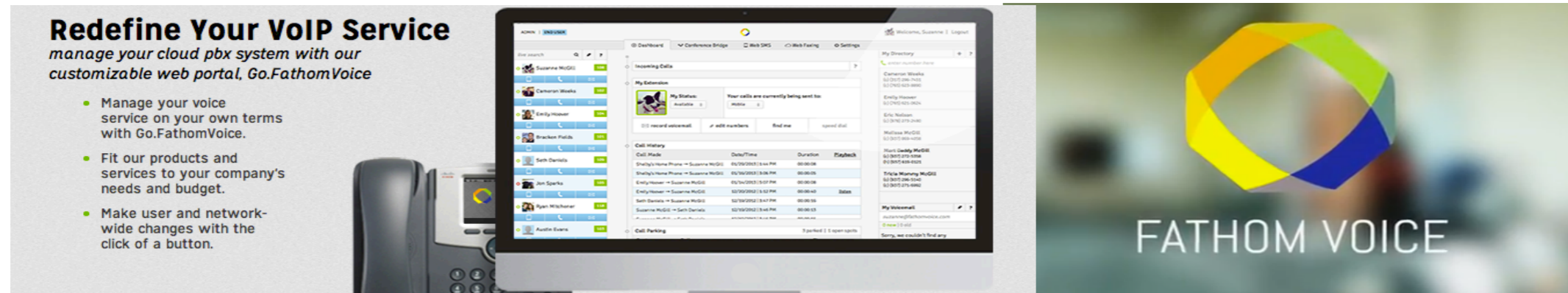


- *Brokers direct local apps to local TEs*
- *TEs maintain local data in memory*
- *All SMs have complete consistent state of database*
- *Commit is synchronous in the local region only*

# Customer - Fathom Voice



# Customer Example - Fathom Voice



Problem	Solution	Benefits
<ul style="list-style-type: none"> <li>• Growing global VoIP customer base; growing latency, data consistency and billing data issues</li> <li>• Competitive differentiation and market expansion</li> <li>• Ability to enhance app hampered by DBMS limitations (MySQL, Amazon RDS)</li> </ul>	<ul style="list-style-type: none"> <li>• Geo-distribute a single logical database</li> <li>• Deliver lower latency, scale out/in easily in the AWS cloud, reduced administration</li> <li>• Greater flexibility in the database, less burden placed in the app</li> </ul>	<ul style="list-style-type: none"> <li>• Cuts latency; presents same data to co-workers in all locations</li> <li>• Makes service faster, more flexible, provides built in HA - all in AWS</li> <li>• Customer can enhance VoIP service without working around DBMS flaws</li> </ul>

# Some Other Examples



# Parting Thoughts

“The more than **50 software makers** that crowded into the red-hot application server market a year ago have **consolidated** and clear leaders are beginning to emerge”

- *Wylie Wong, CNET News, December 1999*

- ▶ 18 Months later there were only a handful of credible offerings
  - ▶ Expect the NextGen database market to consolidate
  - ▶ Expect the winning products to deliver consolidated features, and for the terms NoSQL and NewSQL to go away
  - ▶ Expect distributed transactions to be the central technical challenge
  - ▶ GOOGLE are moving their other applications to GOOGLE FI
- => Expect the leading NextGen databases to evolve towards FI

**Gartner. 2013**  
**CoolVendor**



**NUODB<sup>®</sup>**






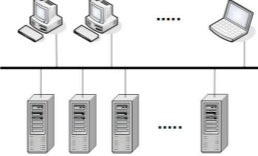
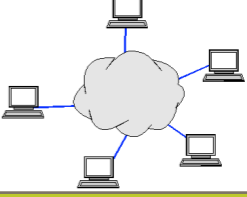


NUODB®

# Database Futures

Barry Morris, NuoDB Inc



# Database History

	Mainframe	Client-Server	Next Gen
Datacenter Architecture			
Lead Vendor			??????
Requirements	<ul style="list-style-type: none"> <li>▶ DB Size: Megabytes</li> <li>▶ #Users: 100's</li> <li>▶ TPS: 10's</li> <li>▶ Latency: seconds</li> <li>▶ Simple Types</li> </ul>	<ul style="list-style-type: none"> <li>▶ DB Size: Gigabytes</li> <li>▶ #Users: 1,000's</li> <li>▶ TPS: 100's</li> <li>▶ Latency: seconds</li> <li>▶ Simple Types, BLOBS</li> <li>▶ SQL</li> <li>▶ ACID Transactions</li> </ul>	<ul style="list-style-type: none"> <li>▶ DB Size: Petabytes</li> <li>▶ #Users: 1,000,000's</li> <li>▶ TPS: 1,000,000's</li> <li>▶ Latency: &lt;0.5s</li> <li>▶ Simple Types, BLOBS, Documents, Media, Geolocation, Time Series etc</li> <li>▶ SQL</li> <li>▶ ACID Transactions</li> <li>▶ Elastic Scalability</li> <li>▶ Mixed Workloads (OLTP/OLAP)</li> <li>▶ 24x7 - zero downtime</li> <li>▶ Active/Active Geodistributed</li> <li>▶ Developer Empowered</li> </ul>



# Forced to Choose?

# NoSQL

On-demand Scale-out

Unstructured Data

Continuous Availability

Geo-distribution

Developer  
Friendly

Powerful Query  
Language

Industry Standards

Data Guarantees

Employee Skills

Existing Data

Tools

# SQL

# Why can't we have it all?

# NoSQL

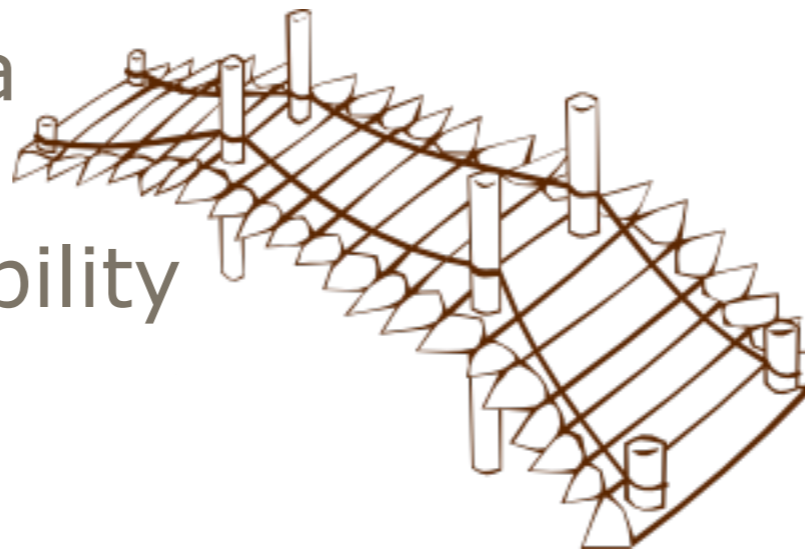
On-demand Scale-out

Unstructured Data

Continuous Availability

Geo-distribution

Developer Friendly



Powerful Query Language

Industry Standards

Data Guarantees X

Employee Skills

Existing Data

Tools

# SQL