SCHOONER
SCALE SMART

How to Maximize Availability, Performance, and Scalability with Synchronous Replication, Auto-Failover, and Flash Optimization

Dr. John R. Busch
Founder and CTO
Schooner Information Technology
John.Busch@SchoonerInfoTech.com

Schooner Information Technology. *October 21, 2011*

Business’ Most Valuable Asset: Its Data

Data

- Most important and valuable component of modern applications and websites

- Driving revolutionary changes in computing and the internet

  - New opportunities for generating revenue

  - More efficient use of current business processes and infrastructure

- Data access downtime or poor performance has a major cost to a business’ bottom line

The Mission-Critical Imperative

“Let me tell you the difference between Facebook and everyone else, we don't crash EVER! If our service is down for even a minute, our entire reputation is irreversibly destroyed

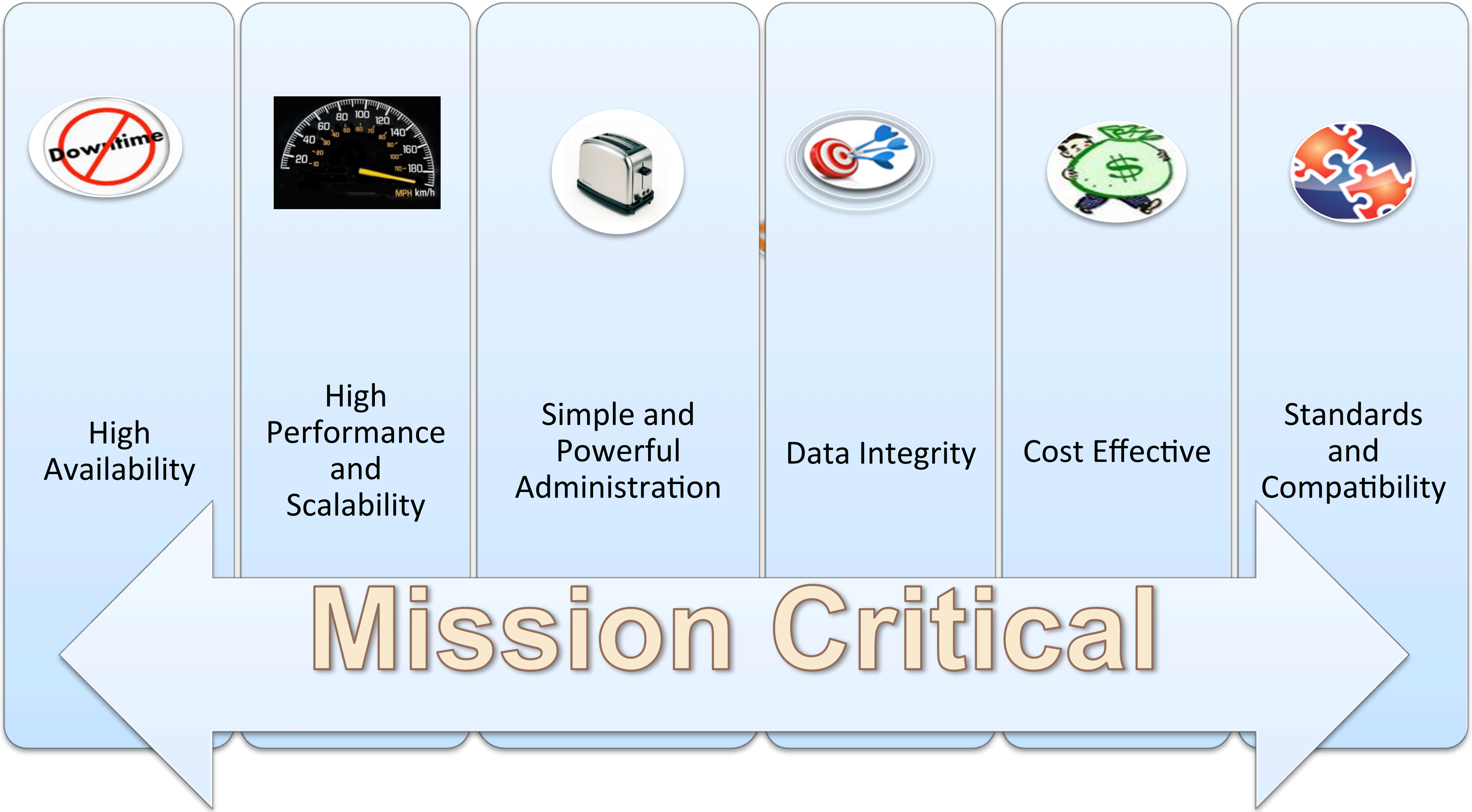
Facebook and Google invest hundreds of millions of dollars every year on custom software and hardware infrastructure to optimize availability, performance, administration, and cost

The Mission-Critical Imperative

- Providing high data availability, excellent response time is critical for key classes of businesses
  - Web 2.0
  - eCommerce
  - High-volume websites
  - Telecommunications

- They require a mission critical database

Mission-Critical Database Requirements



High Availability

High Performance and Scalability

Simple and Powerful Administration

Data Integrity

Cost Effective

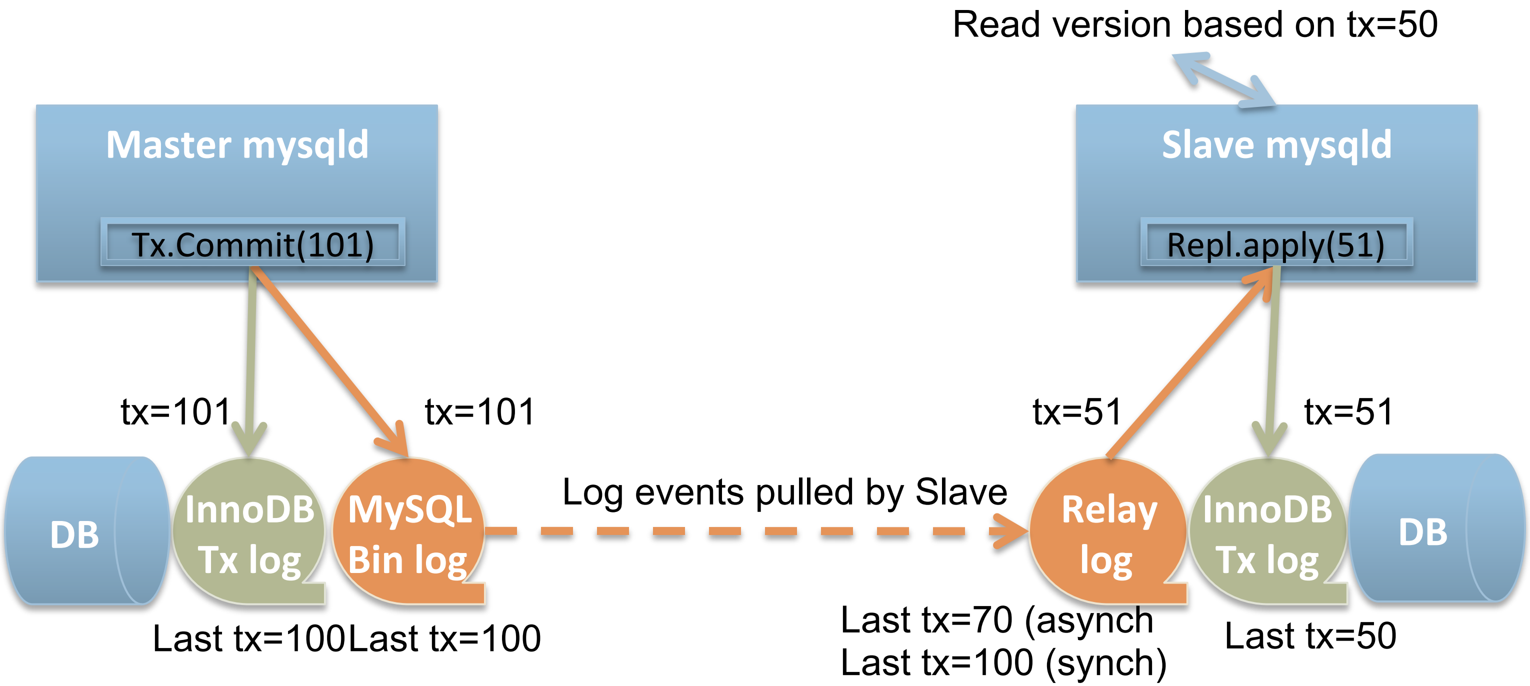
Standards and Compatibility

Mission Critical

Mission-Critical Database Goals and Metrics

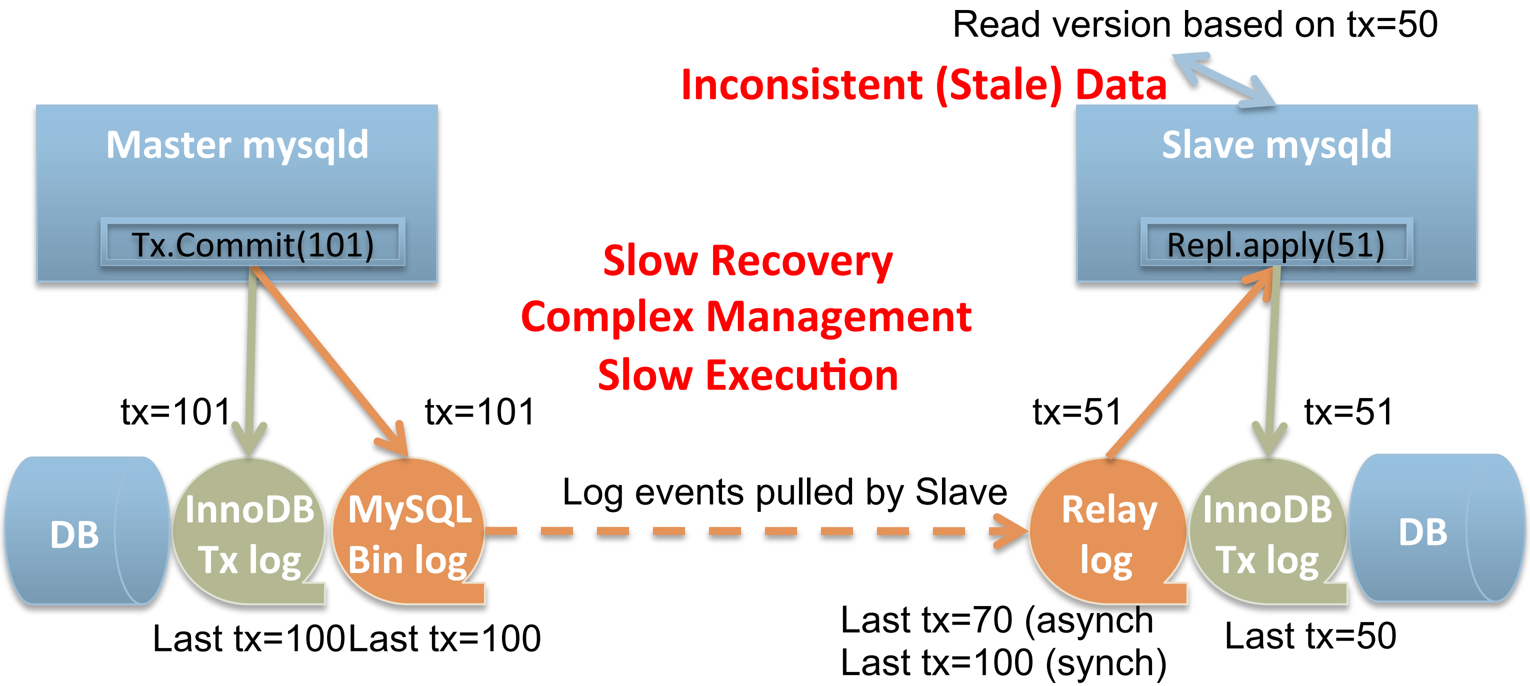
Goals	Metrics
High Availability	Service unavailability (minutes/year) from failures, disaster recovery, or during planned administration
High Data Integrity	Probability of data loss or corruption; data consistency levels
High Performance and Scalability	Transaction throughput, response time; performance scalability; performance stability
Simple and powerful administration	Ease of cluster administration; fail-over automation; monitoring and optimization tools
Cost effective	Total cost of ownership (TCO); return on investment (ROI)
Standards and Compatibility	Level of standards compliance and certification

Loosely-Coupled Asynchronous and Semi-Synchronous Replication



Example Products : MySQL Enterprise 5.1 Asynchronous and 5.5/5.6 Semi-Synchronous Replication

Loosely-Coupled Asynchronous and Semi-Synchronous Replication



Potential Data Loss

Example Products : MySQL Enterprise 5.1 Asynchronous and 5.5/5.6 Semi-Synchronous Replication

Loosely-Coupled Asynchronous and Semi-Synchronous Replication

Limited Service Availability

- Master fail-over, re-synch of slaves

Limited Data Integrity

- Lost data; inconsistent Data

Limited Performance and Utilization

- Low throughput and low utilization

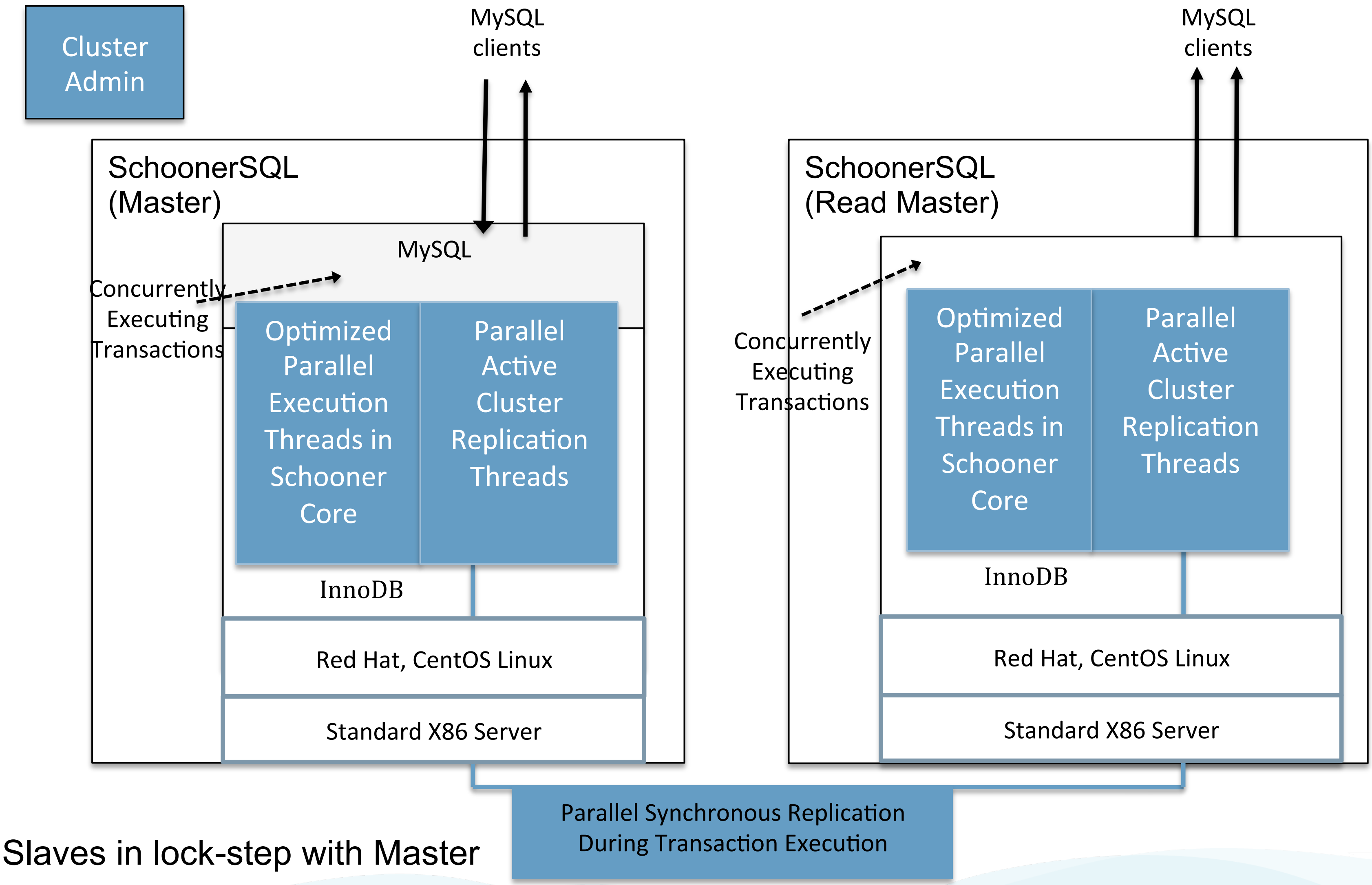
Complex Administration

- Manual processes, slave re-synch

High Cost of Ownership

- High capital expense from server sprawl
- Increased operating expense from power, space, admin
- Reduced revenue and customer satisfaction from service downtime

Tight Coupling and Fully Synchronous Replication



Cluster Admin

MySQL clients

SchoonerSQL (Master)

MySQL

Concurrently Executing Transactions

Optimized Parallel Execution Threads in Schooner Core

Parallel Active Cluster Replication Threads

InnoDB

Red Hat, CentOS Linux

Standard X86 Server

MySQL clients

SchoonerSQL (Read Master)

Concurrently Executing Transactions

Optimized Parallel Execution Threads in Schooner Core

Parallel Active Cluster Replication Threads

InnoDB

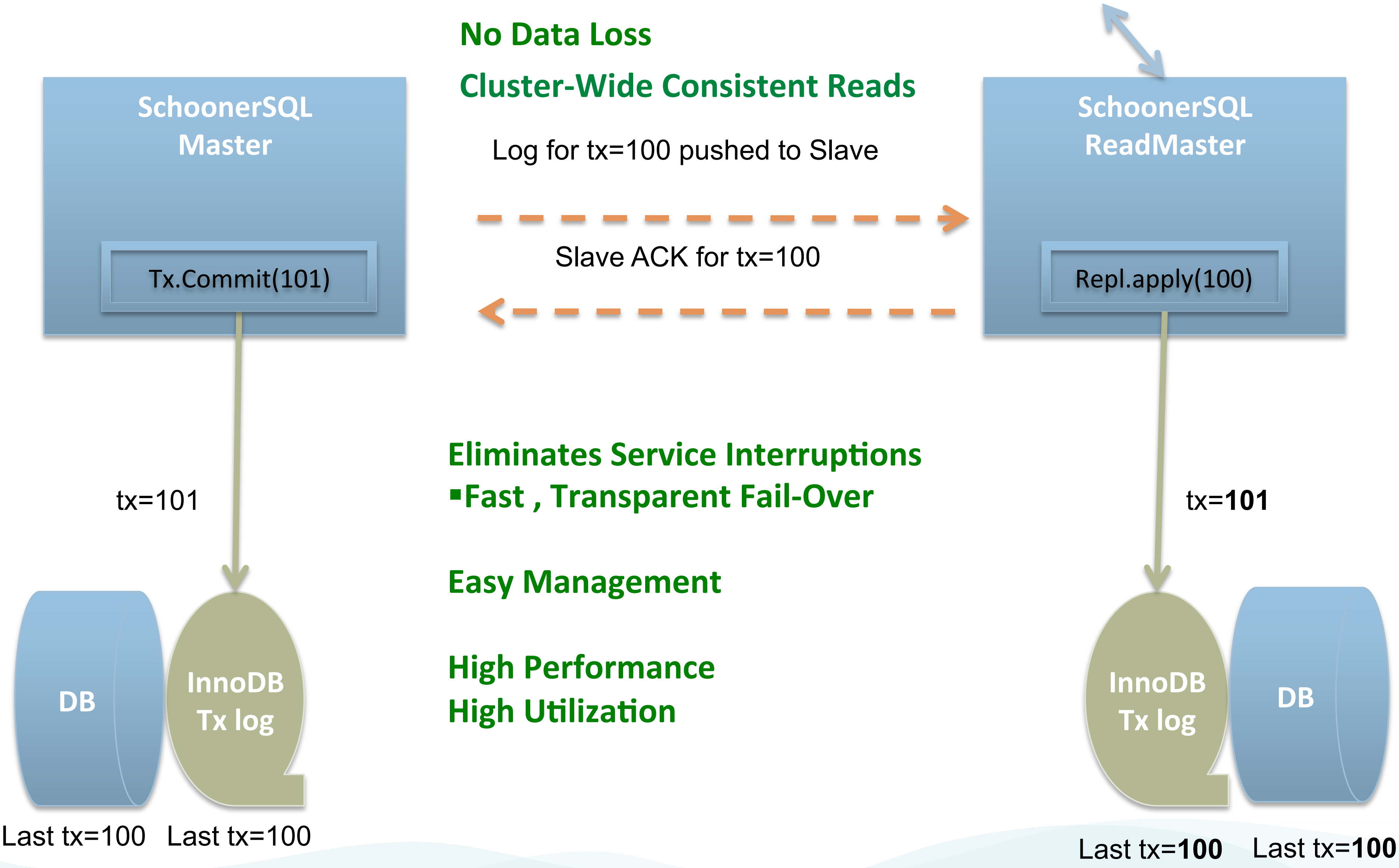
Red Hat, CentOS Linux

Standard X86 Server

Parallel Synchronous Replication During Transaction Execution

- Slaves in lock-step with Master
- At master transaction commit, all Slaves guaranteed to have received and committed the changes

Tight-Coupling and Synchronous Replication



No Data Loss

Cluster-Wide Consistent Reads

SchoonerSQL Master

Tx.Commit(101)

Log for tx=100 pushed to Slave

Slave ACK for tx=100

SchoonerSQL ReadMaster

Repl.apply(100)

tx=101

tx=**101**

DB

InnoDB Tx log

Last tx=100 Last tx=100

Last tx=**100** Last tx=**100**

Eliminates Service Interruptions

- **Fast , Transparent Fail-Over**

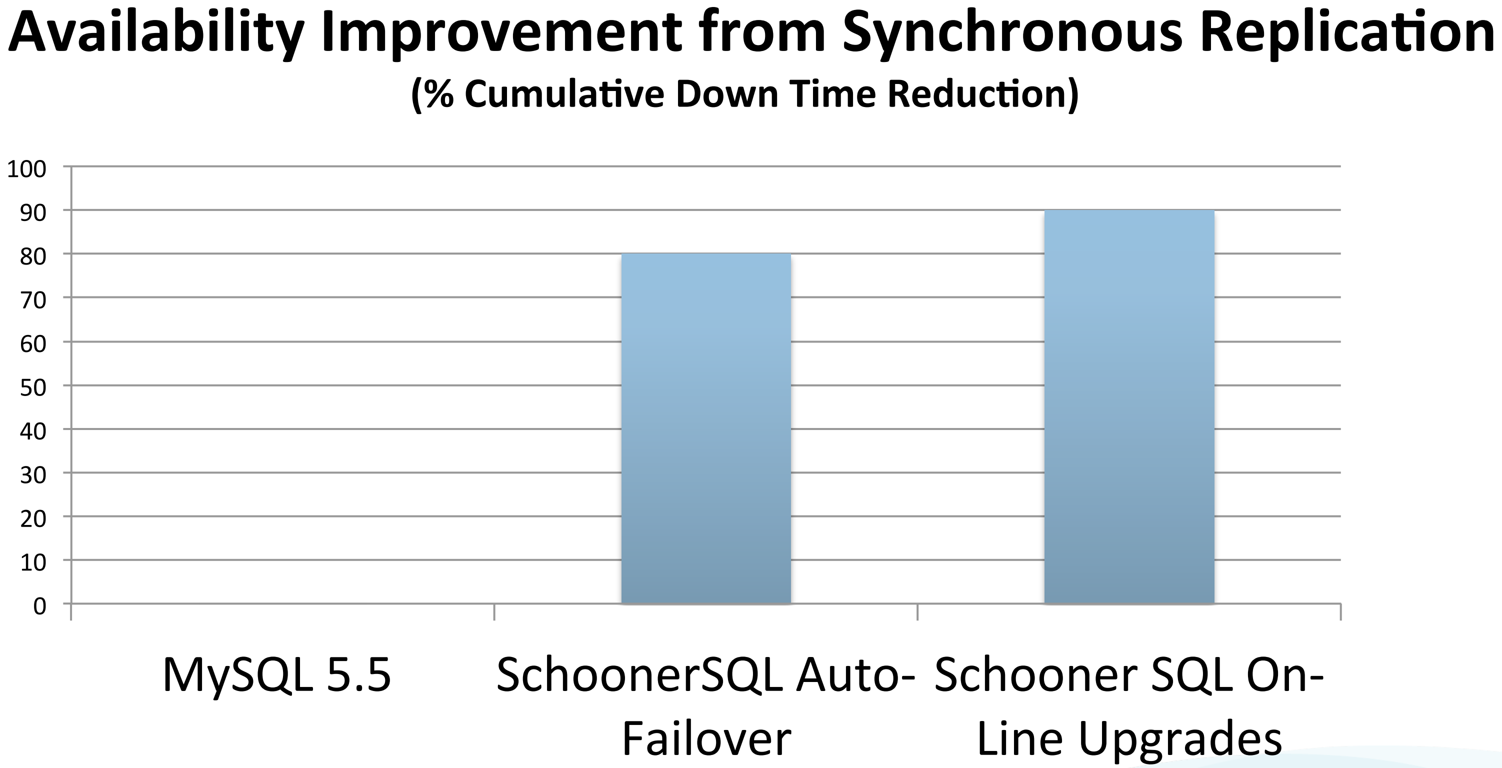
Easy Management

High Performance

High Utilization

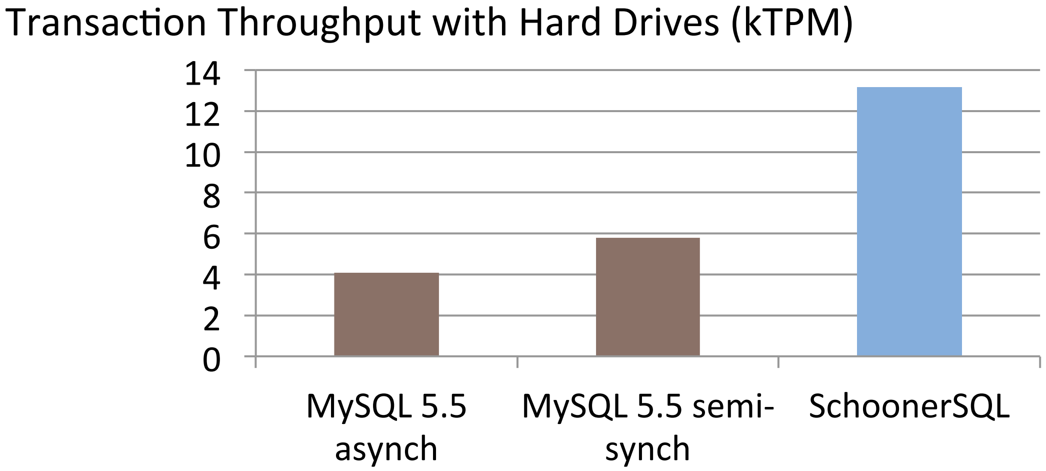
Tight Coupling and Synchronous Replication Can Improve Service Availability by 90%

Tightly-coupled MySQL synchronous replication can provide much higher service availability than that achievable with asynchronous or semi-synchronous replication



Tight Coupling and Synchronous Replication Can Provide Much Higher Performance Throughput per Server

Synchronous Transaction Throughput per Server can be Much greater Than Asynchronous or Semi-Synchronous (with hard disc drives (HDDs))



Measurement Configuration
- 2 node Master-Slave configuration
- 2 socket Westmere
- 72GB DRAM

DBT2 open-source OLTP version of TPC-C
- 1000 warehouses, 32 connections
- 0 think-time
- Result metric: TPM (new order)

Tight Coupling and Synchronous Replication Can Scale Vertically with Commodity Flash Memory, Cores

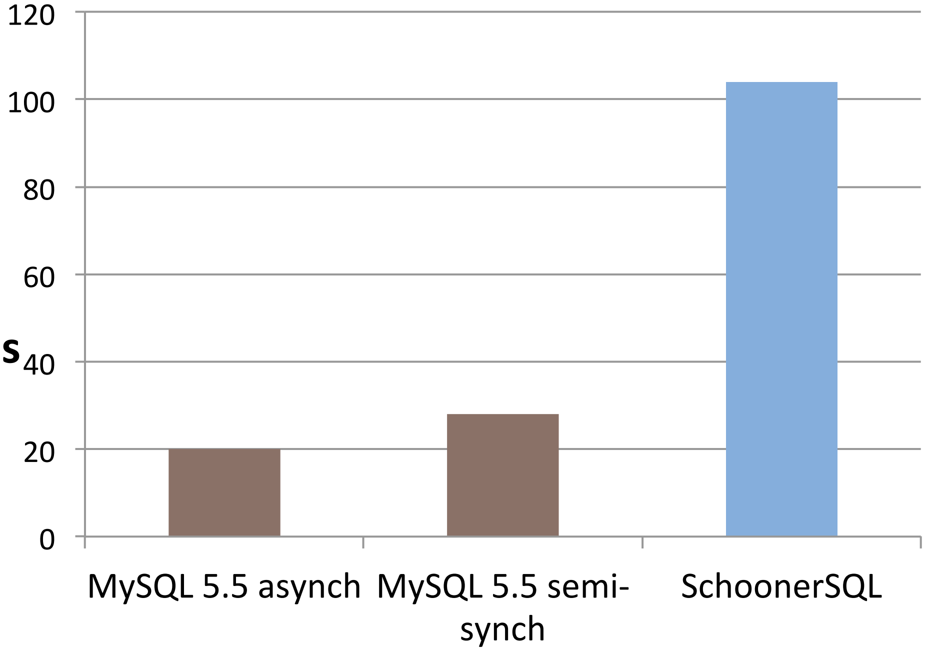
DBT2 open-source OLTP version of TPC-C

- 1000 warehouses, 32 connections
- 0 think-time
- Result metric: TPM (new order)

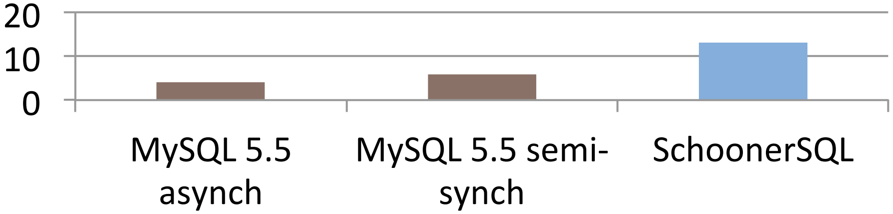
Measurement Configuration

- 2 node Master-Slave configuration
- 2 socket Westmere
- 72GB DRAM

Transaction Throughput with Flash Drives

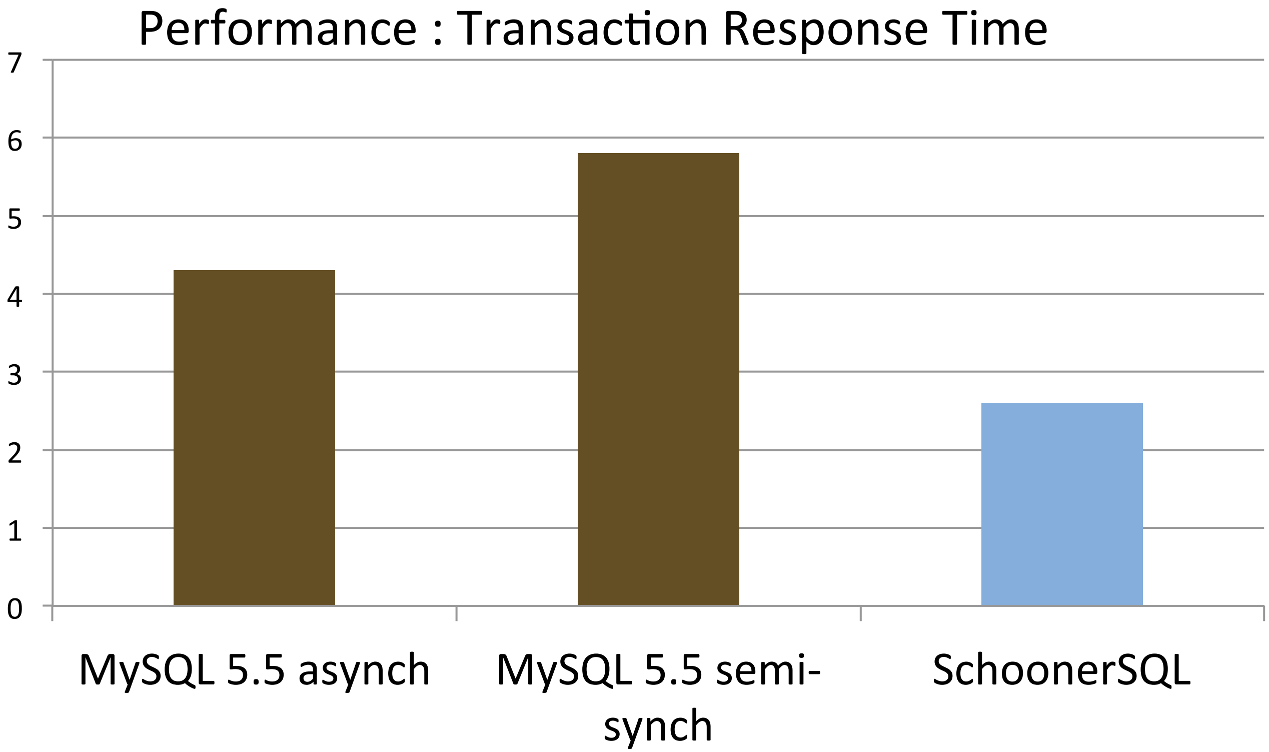


Transaction Throughput with Hard Disc Drives



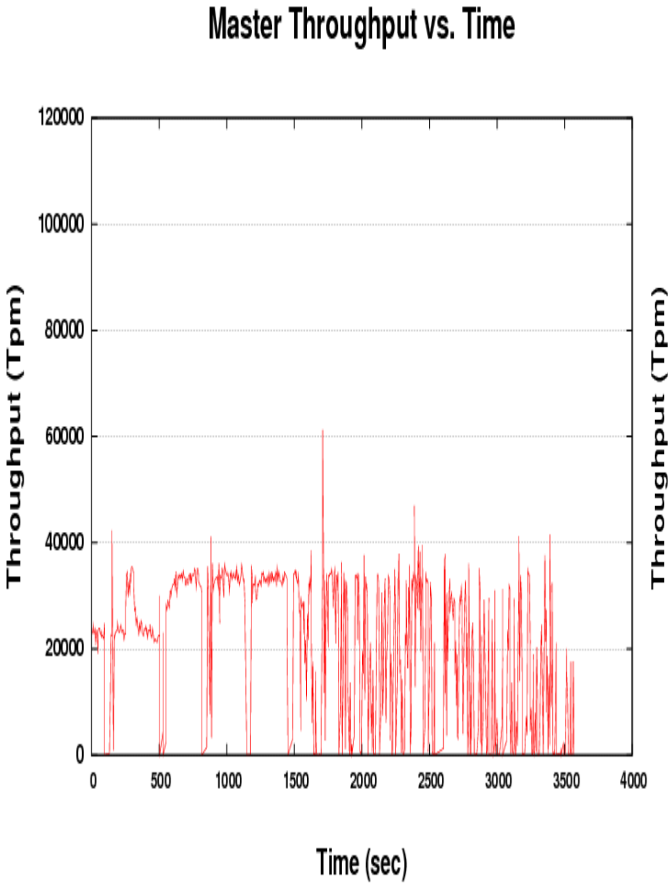
Tight Coupling and Synchronous Replication Can Lower Response Times

Response Time (ms)

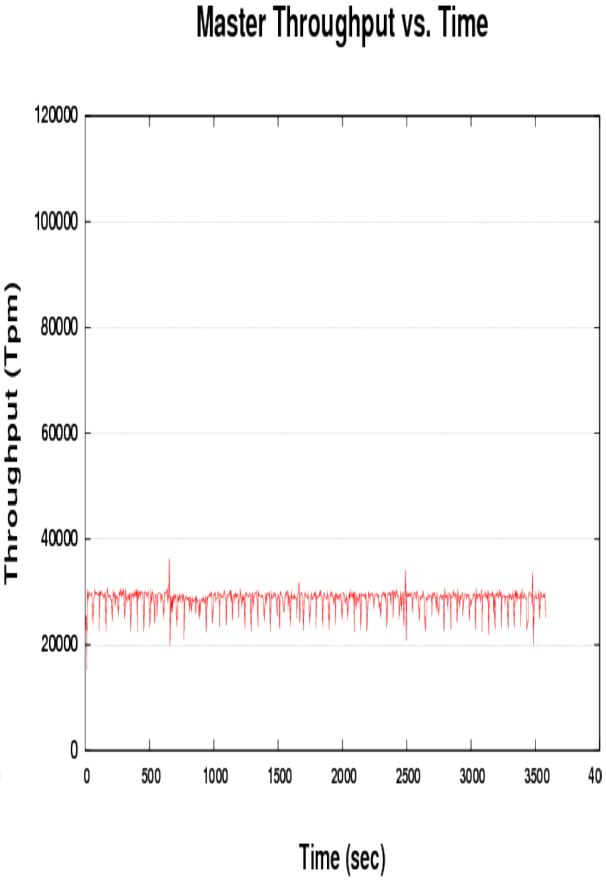


Tight Coupling and Synchronous Replication Can Provide Higher Performance Stability

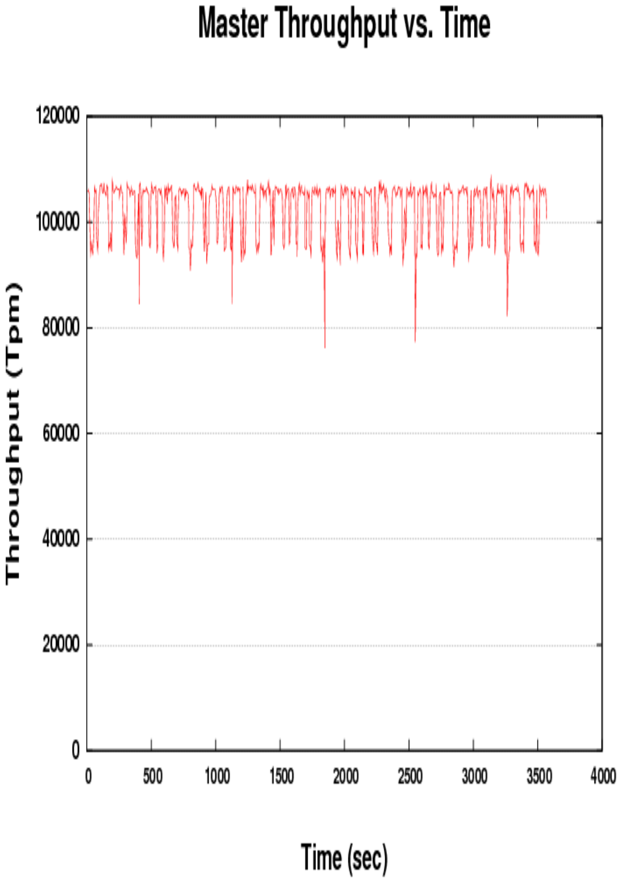
MySQL 5.5 Asynchronous



MySQL 5.5 Semi-synchronous



SchoonerSQL

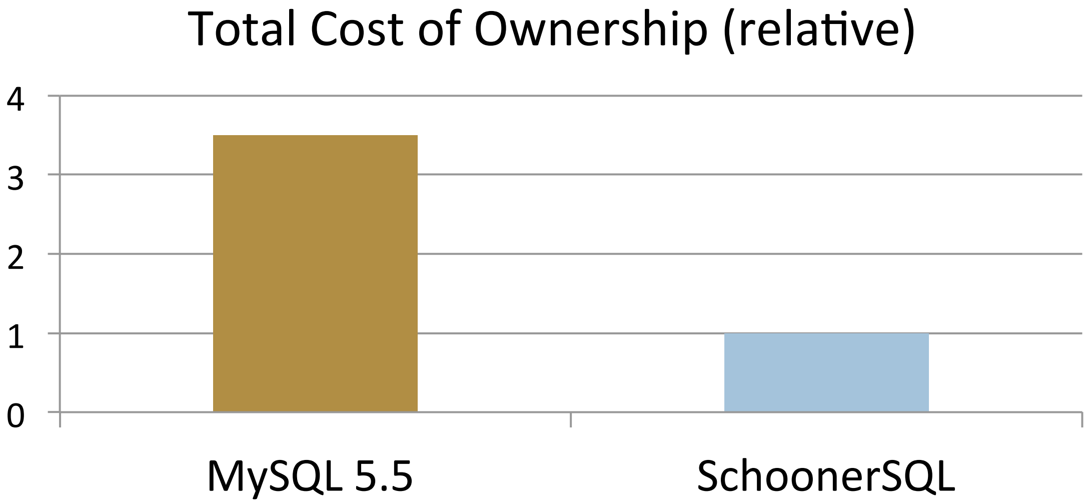


Tight Coupling and Synchronous Replication Can Lower Total Cost of Ownership

Lower Cost

▪Reduced capital and operating costs through reduction in servers, power, space, admin

▪Savings from increased service availability and associated revenue and customer retention

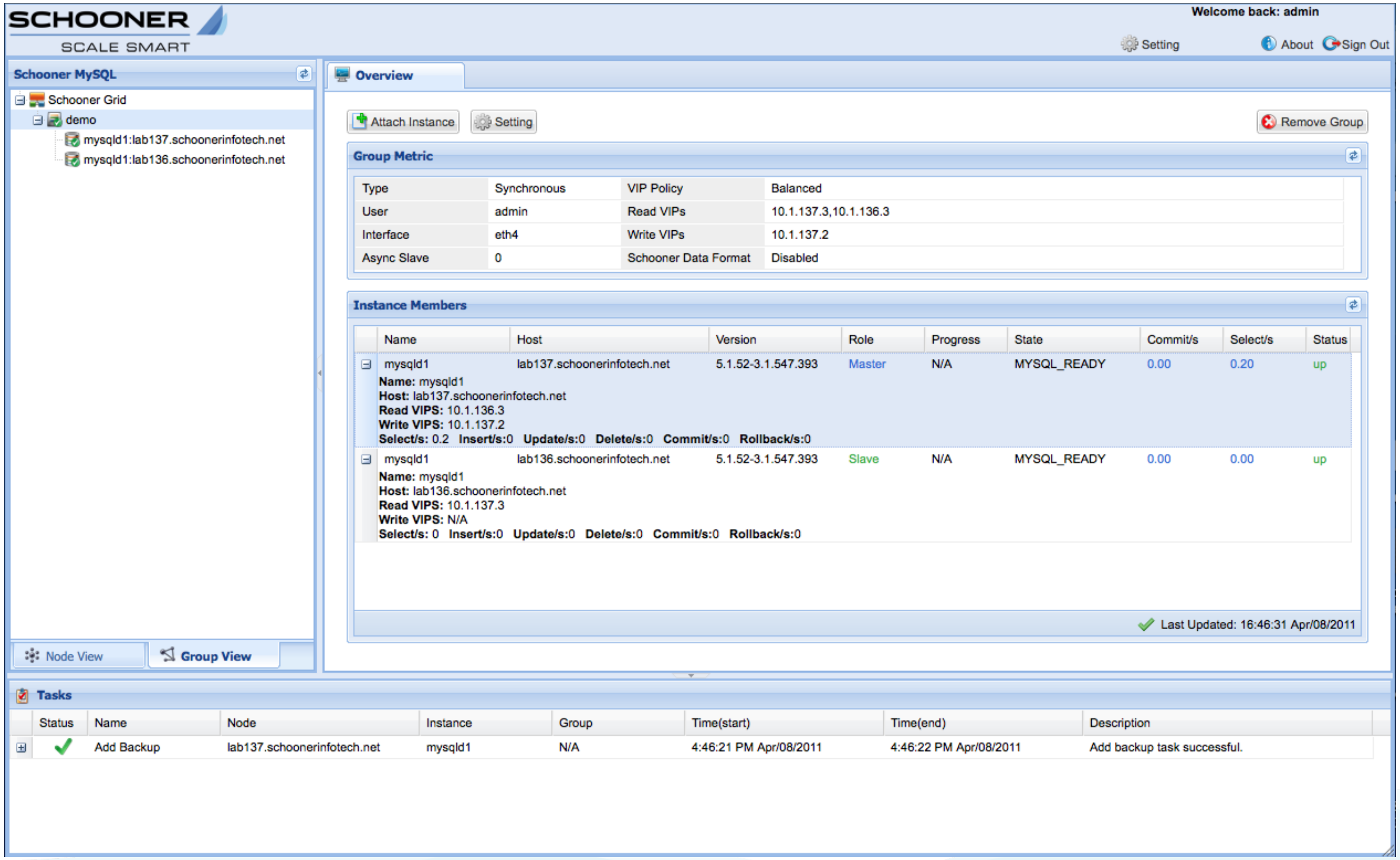


▪TCO and ROI models are customer and workload specific

▪Function (throughput/server; server, rack, and network costs, software license and support costs, admin costs; space and power costs; cost of downtime)

Tight Coupling and Synchronous Replication Can Simplify Administration

- **Fail-over can be completely automatic and instant**
  - requiring no administrator intervention or service interruption
- **Cluster Administrator GUI and CLI can provide a single point for cluster-wide management**
  - single click slave creation and database migration



Group Metric

Type	Synchronous	VIP Policy	Balanced
User	admin	Read VIPs	10.1.137.3,10.1.136.3
Interface	eth4	Write VIPs	10.1.137.2
Async Slave	0	Schooner Data Format	Disabled

Instance Members

Name	Host	Version	Role	Progress	State	Commit/s	Select/s	Status
mysqld1	lab137.schoonerinfotech.net	5.1.52-3.1.547.393	Master	N/A	MYSQL_READY	0.00	0.20	up
mysqld1	lab136.schoonerinfotech.net	5.1.52-3.1.547.393	Slave	N/A	MYSQL_READY	0.00	0.00	up

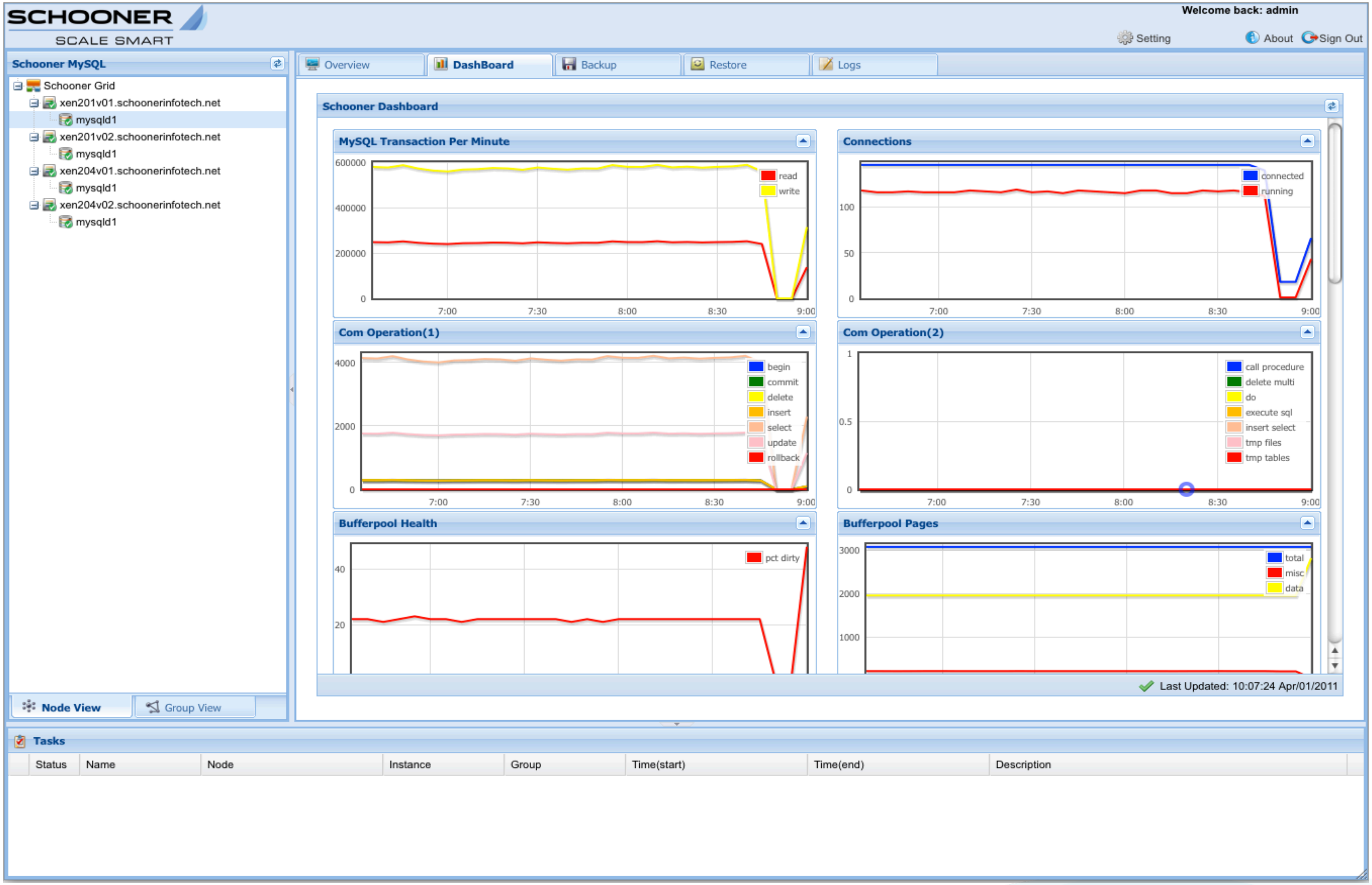
Name: mysqld1
Host: lab137.schoonerinfotech.net
Read VIPS: 10.1.136.3
Write VIPS: 10.1.137.2
Select/s: 0.2 **Insert/s:**0 **Update/s:**0 **Delete/s:**0 **Commit/s:**0 **Rollback/s:**0

Name: mysqld1
Host: lab136.schoonerinfotech.net
Read VIPS: 10.1.137.3
Write VIPS: N/A
Select/s: 0 **Insert/s:**0 **Update/s:**0 **Delete/s:**0 **Commit/s:**0 **Rollback/s:**0

✓ Last Updated: 16:46:31 Apr/08/2011

Status	Name	Node	Instance	Group	Time(start)	Time(end)	Description
✓	Add Backup	lab137.schoonerinfotech.net	mysqld1	N/A	4:46:21 PM Apr/08/2011	4:46:22 PM Apr/08/2011	Add backup task successful.

Extend with Powerful Administration : Monitoring, trouble-shooting, tuning

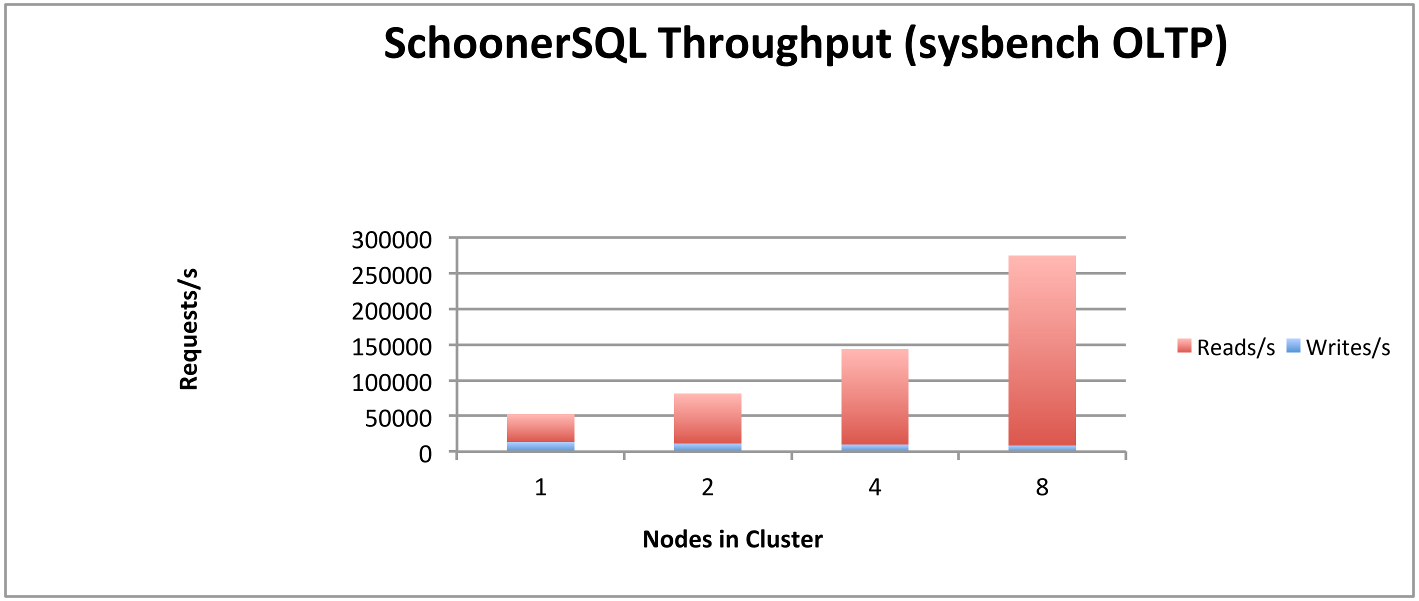


Welcome back: admin

✓ Last Updated: 10:07:24 Apr/01/2011

Status	Name	Node	Instance	Group	Time(start)	Time(end)	Description

Unlimited Query Scaling

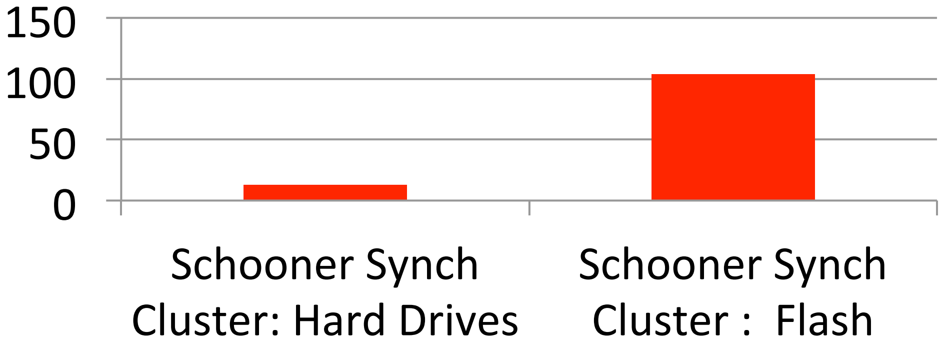


Query Scaling in a Synchronous Replication Group

- Fully replicated Master/Slave cluster
  - No cluster overhead for adding queries to a slave
  - Can add synchronous query nodes linearly
- With partitioned databases, scaling is sub-linear with severe cross node query degradation

Scaling Updates

- Database Update Scalability
  - Vertically scale with commodity : flash memory, more cores, higher frequency



- Compelling option exploiting low cost, high performance commodity technology

Unlimited Update Scaling : Optimized Transparent Sharding

- Database Update Scalability

  …After Optimal Vertical Scaling:

  Horizontally Scale Through Partitioning (Sharding)

  - Database workload aware
    - Administrator analysis and configuration tools
    - allows layout and query data access optimization

  - Application Transparent
    - Dynamic query execution across shards

Extending Synchronous Replication:
WAN Replication and Disaster Recovery

- WAN/geographically dispersed data centers
  - Requires Asynchronous replication
    - Can't add additional ~100ms with high potential variance to query response time for synchronous replication

- HA Requirement: WAN asynch slave should automatically fail-over when synchronous master fail-over occurs
  - WAN asynchronous replication must be integrated with synch replication group

- Data Integrity Requirement : Remote consistency lag and recovery time should be ~ WAN latency
    - Maximize WAN data consistency
    - Minimize disaster recovery time
  - Requires high performance asynchronous replication
    - Must have multi-threaded asynchronous parallelizing updates

Comparison of Alternatives

FEATURES & BENEFITS	MYSQL 5.5	DRBD	ScaleDB	MYSQL NDB CLUSTER	CONTINUENT (TUNGSTEN)	CLUSTRIX	SCHOONER SQL
Synchronous Replication for InnoDB (Guaranteed Data Consistency)	No	Limited	No	No	No	No	Yes
# Node Failures before Service Downtime (Failure Resistance)	Two	Two	Three	Four	Two	Two	Eight
Eliminates Slave Lag (100% Data Consistency and Zero Data Loss)	No	No	N/A	N/A	No	N/A	Yes
Automated Fail-Over (LAN/MAN/WAN)	No	No	No	No	No	No	Yes
Performance Across WAN	Low	Low	Low	Low	Low	Low	High
Full & Incremental Online Backup Integrated with GUI (Zero Downtime)	Limited	No	No	No	No	No	Yes
Online Software & Hardware Upgrades (Zero Downtime)	No	No	No	Low	No	Low	High
Elastic Cluster (add or remove nodes with ease - Zero Downtime)	No	No	Medium	Medium	Low	Medium	High
Performance with Flash Memory	Low	Low	Low	Low	Low	Medium	High
Cost (TCO)	Medium	High	High	High	High	High	Low

SchoonerSQL - Come Visit Our Booth and China Team

MISSION CRITICAL

Highest Availability

- No service interruption for planned or unplanned database downtime
- Instant automatic fail-over
- On-line upgrade and migration
- 90% less downtime vs. MySQL 5.5
- Full WAN support with master auto-failover

Highest Performance and Scalability

- 4-20x more throughput/server vs. MySQL 5.5
- High performance synchronous and asynchror replication

Compelling Economics

- Cut server capex (consolidation)
- Cut opex (power, pipe, DBA time)
- Increase revenue (eliminate service interruptions)
- TCO 70% cheaper than MySQL 5.5

100% MySQL Enterprise InnoDB Compatible

Highest Data Integrity

- No lost data
- Cluster-wide data consistency

Visibility and Control

- Easy cluster administration
- No error-prone manual processes
- Monitoring and Optimization

Out-of-the-box Product

- Full MySQL + InnoDB: not a toolkit
- Free your staff to build your business, not a custom database

Broad Industry Deployment

- eCommerce, Social Media, Telco, Financial Services, Education
- High volume web sites
- Geographically distributed websites

ebay Comcast 37signals OSL iStockphoto XOOM gutefrage.net HOLTZBRINCK DIGITAL

Evaluating the Options and Trade-offs for Your Data Center? Let Schooner Help!

CONTACT SCHOONER

Schooner Information Technology, Inc.
501 Macara Avenue, Suite 101
Sunnyvale, CA 94085 USA
Tel: +1 408-773-7500
www.schoonerinfotech.com
Email: info@schoonerinfotech.com

Schooner中国
地址：杭州市西湖区教工路23号百脑汇大厦18楼
传真：057189731509　　电话：057189731653
销售电话：13867476875
Email: salescn@schoonerinfotech.com

Thank You!

QCon

- 北京站 · 2012年4月18~20日
  www.qconbeijing.com（11月启动）
- QCon杭州站官网和资料
  www.qconhangzhou.com

全球企业开发大会

INTERNATIONAL
SOFTWARE DEVELOPMENT
CONFERENCE