QCon 全球软件开发大会
INTERNATIONAL SOFTWARE DEVELOPMENT CONFERENCE

BEIJING 2017

MySQL協助您搭建全方位的高可用應用

杜修文
甲骨文全球事業部

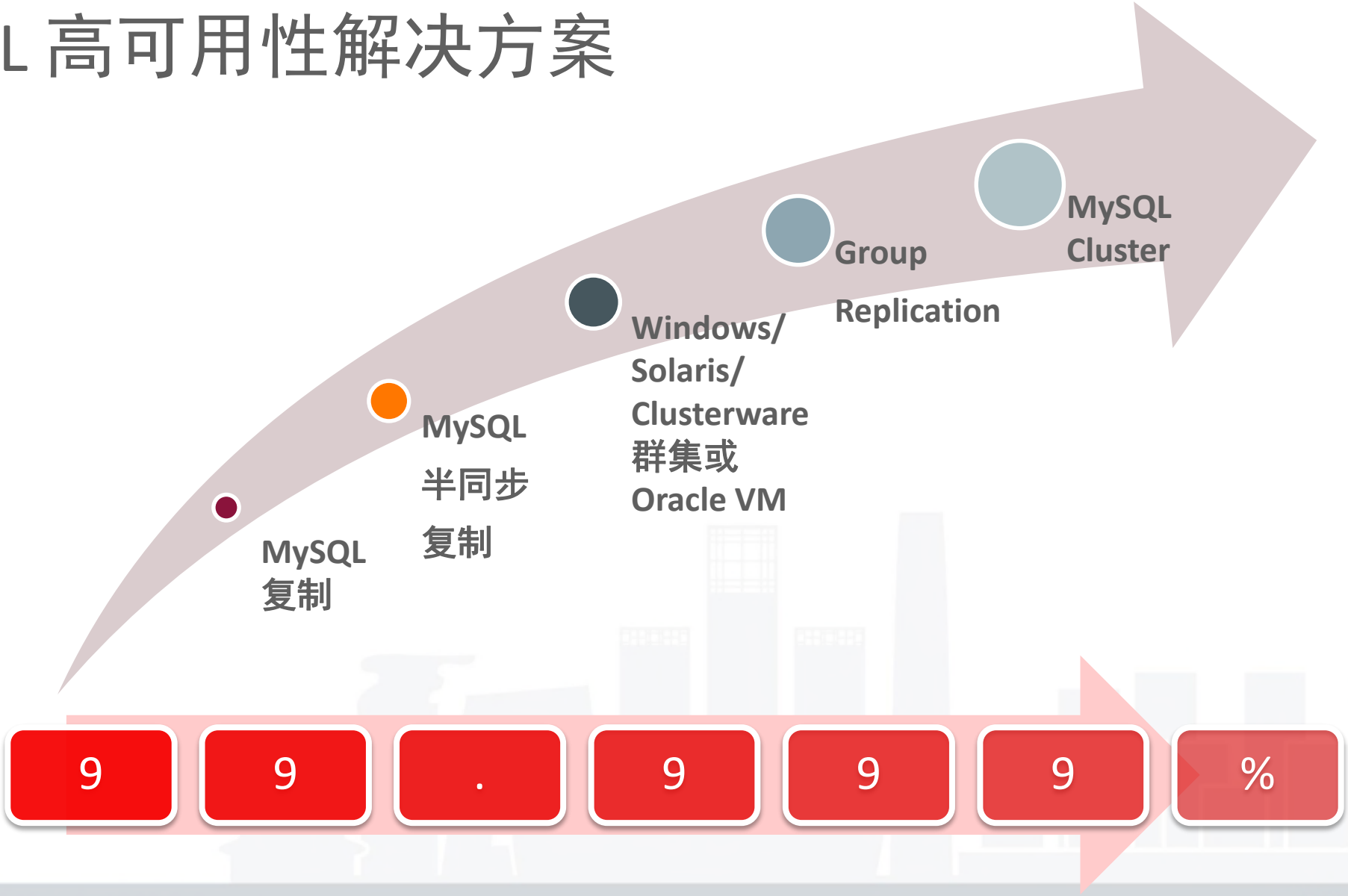
主办方 Geekbang 极客邦科技 InfoQ

安全港声明

以下内容旨在阐明产品的整体方向。该内容仅供参考，不可纳入任何合同。该信息不承诺提供任何资料、代码或功能，并且不应该作为制定购买决策的依据。本文档所述的 Oracle 产品的任何特性或功能的开发、发行和时间规划均由 Oracle 自行决定。

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

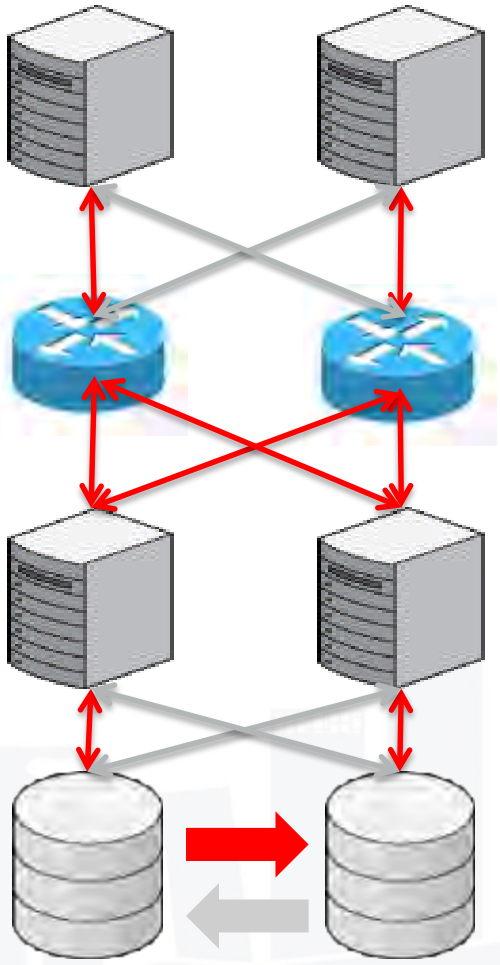
MySQL 高可用性解决方案



版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

支持高可用性的各个层

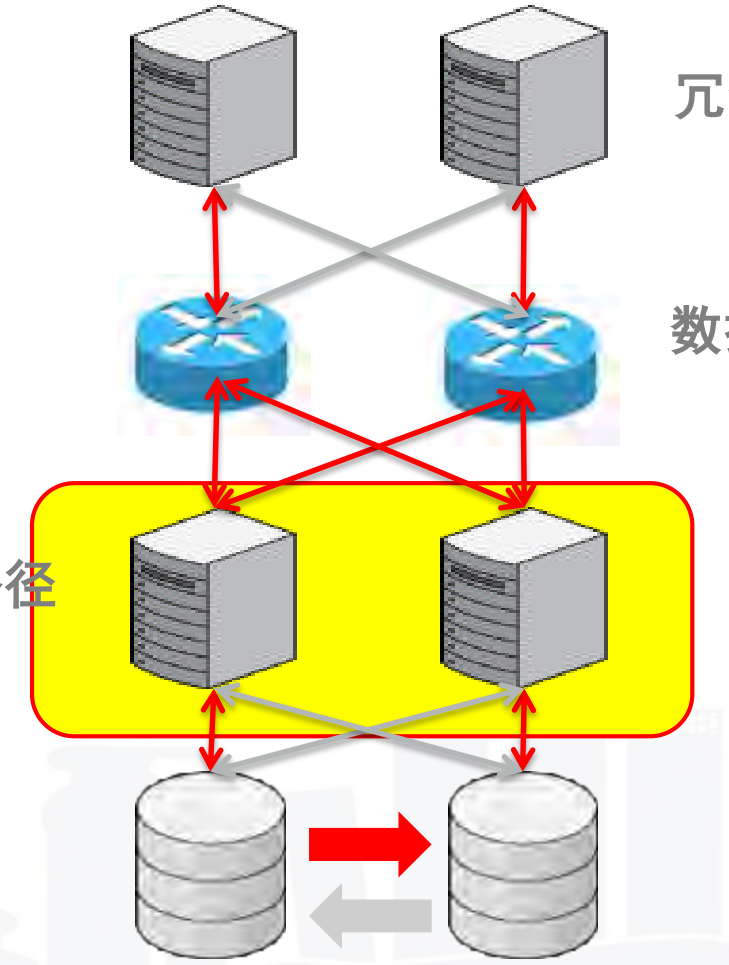
并不仅仅是可靠地存储数据



版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

支持高可用性的各个层

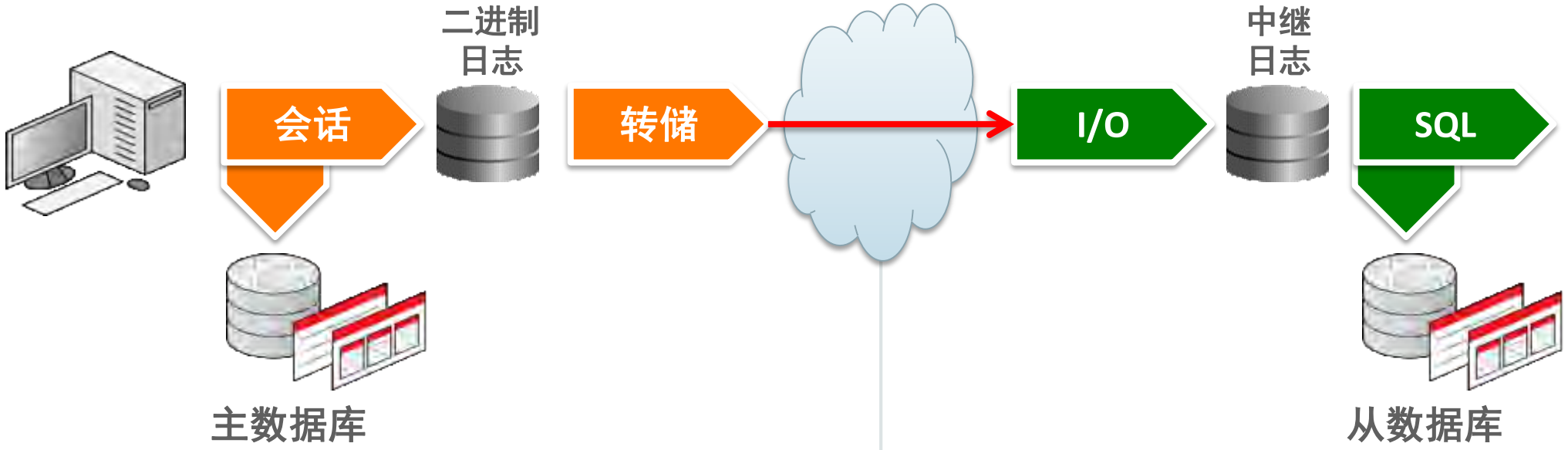
并不仅仅是可靠地存储数据



- **MySQL Replication**
- Shared Storage
- Group Replication
- MySQL Cluster

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

MySQL 复制工作流

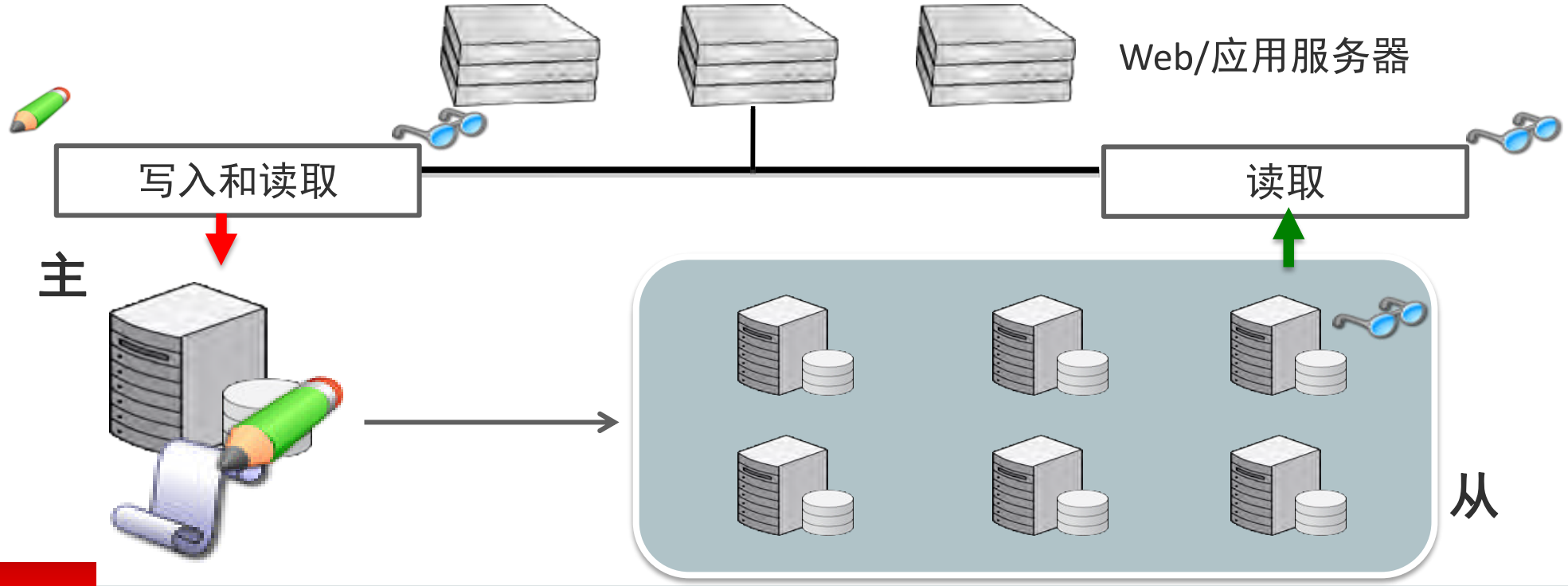


- 会话线程：处理来自应用程序的查询 — 将数据写入主数据库，将关联事件写入二进制日志
- 转储线程：读取二进制日志中的事件，然后将其发送到从数据库
- I/O 线程：接收复制事件，并将其存储在从数据库的中继日志中
- SQL 线程：读取从数据库的中继日志中的复制事件，然后将其应用到从数据库

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

为何进行复制？

- 将数据库从“主服务器”复制到“从服务器”
  - 数据的冗余副本奠定了高可用性的基础
  - 通过在复制场中进行分布式查询来扩展



版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

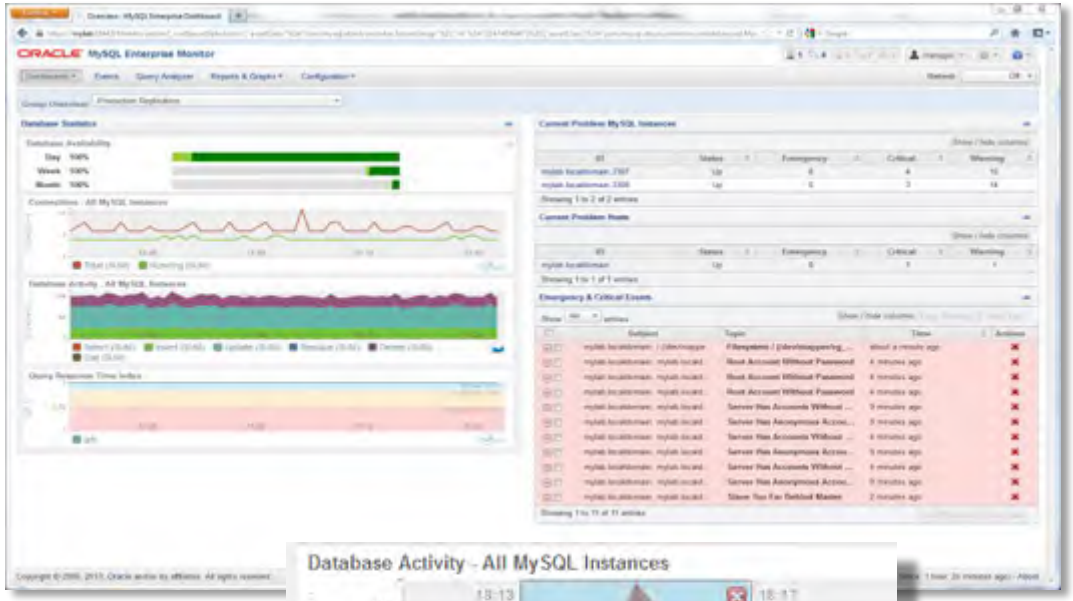
异步复制与同步复制

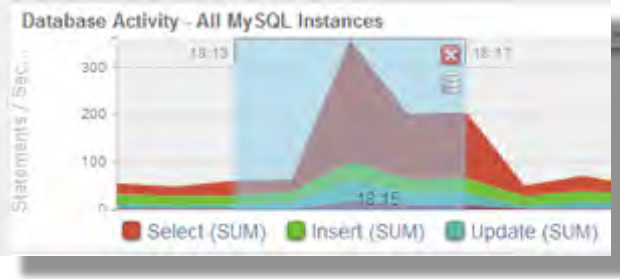
- 异步
  - MySQL 默认设置
  - **并行**：主数据库向应用程序发送确认，在存储引擎上提交，然后将事务发送到从数据库
    - 快速
    - 如果主数据库停机有可能丢失更改
- 半同步
  - MySQL 5.5+ — 在 MySQL 5.7 中进行了增强
  - **串行**：主数据库等待从数据库记录更改并将更改刷新到磁盘，然后以**并行方式**提交到存储引擎并向应用程序发送确认
    - 中度延迟
    - 无损 (MySQL 5.7)
- 同步
  - 仅 MySQL Cluster 提供
  - **串行**：主数据库等待更改应用于所有从数据库后，再向应用程序发送确认
    - 较长的延迟
    - 如果采用主动/主动模式，则最适合小型事务处理
    - 无损

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

MySQL Enterprise Monitor

- 在 10 分钟内开始监视 MySQL
- 实时监视 MySQL 性能和可用性
- 直观地查找和修复有问题的查询
- 监视磁盘，以便制定容量规划
- 适合部署到云的体系结构
  - 不需要代理
- 可选的代理方式能提供高级主机/OS 监视





Current	Worst	Subject	Topic
		mylab.localdomain, mylab.localdomain:3306	Root Account Without Password
		mylab.localdomain, mylab.localdomain:3306	Server Has Accounts Without A Password
		mylab.localdomain, mylab.localdomain:3306	Average Statement Execution Time Excess...
		mylab.localdomain, mylab.localdomain:3306	SQL Statement Generates Errors or Warnings
		mylab.localdomain, mylab.localdomain:3306	Server Has Anonymous Accounts
		mylab.localdomain, mylab.localdomain:3306	MySQL Instance Is Experiencing A Query P...
		mylab.localdomain, mylab.localdomain:3306	InnoDB Log Buffer Flushed To Disk After Ea...
		mylab.localdomain, mylab.localdomain:3306	User Has Rights To Database That Does Not...

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

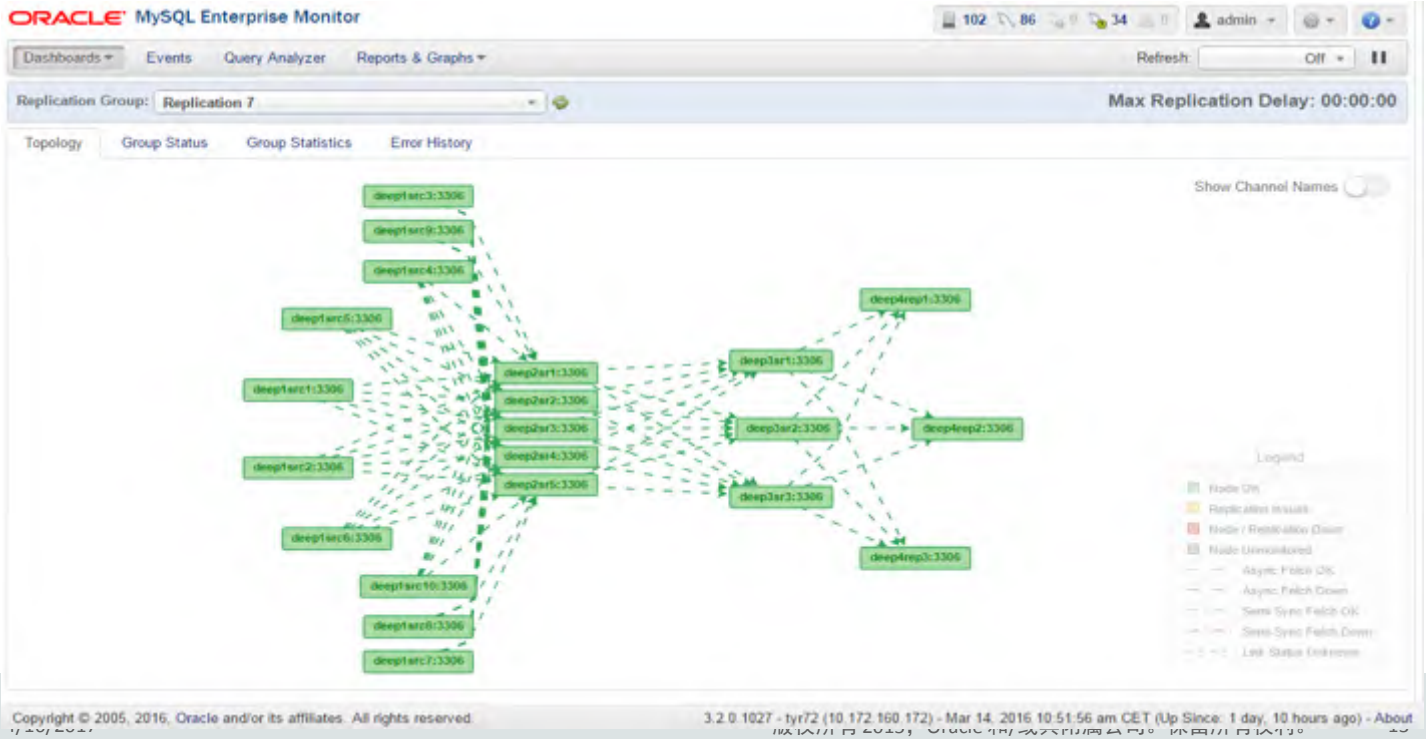
Enterprise Replication Monitor

- 自动搜索复制拓扑
- 主/从性能监视
- 复制指导
- 有关最佳做法的复制建议

“我每天使用 MySQL Enterprise Monitor 监视和跟踪 MySQL 数据库。该软件提供了快速的一站式服务，可用来密切监视数据库。”

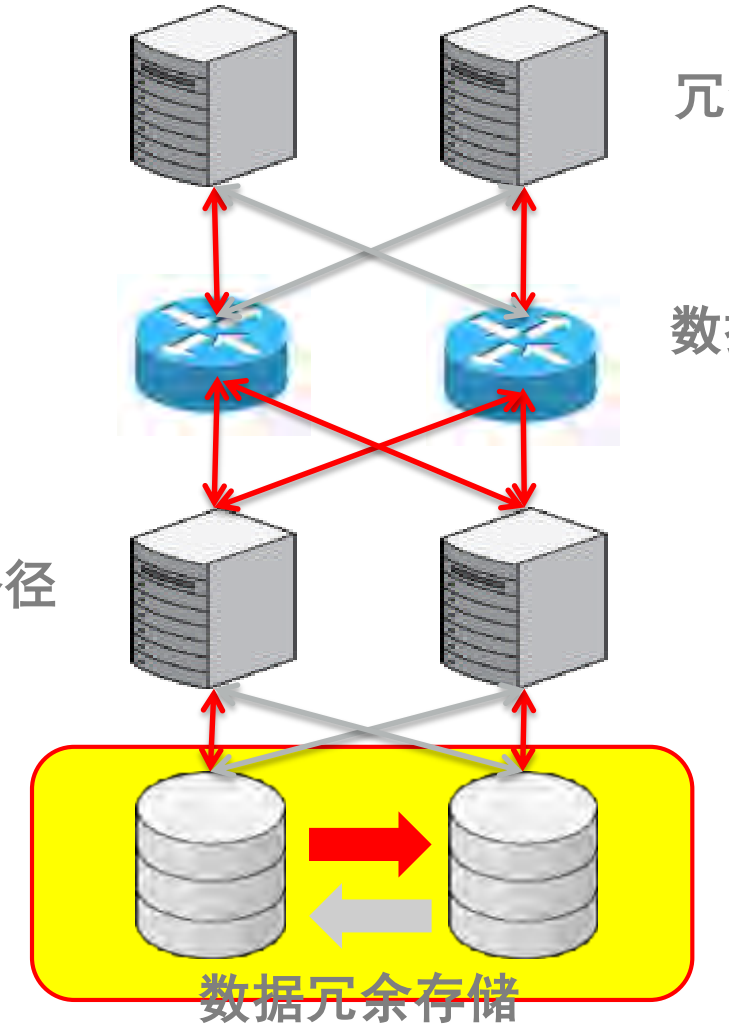
-Wes Homer，
高级系统和网络管理员

InsideView



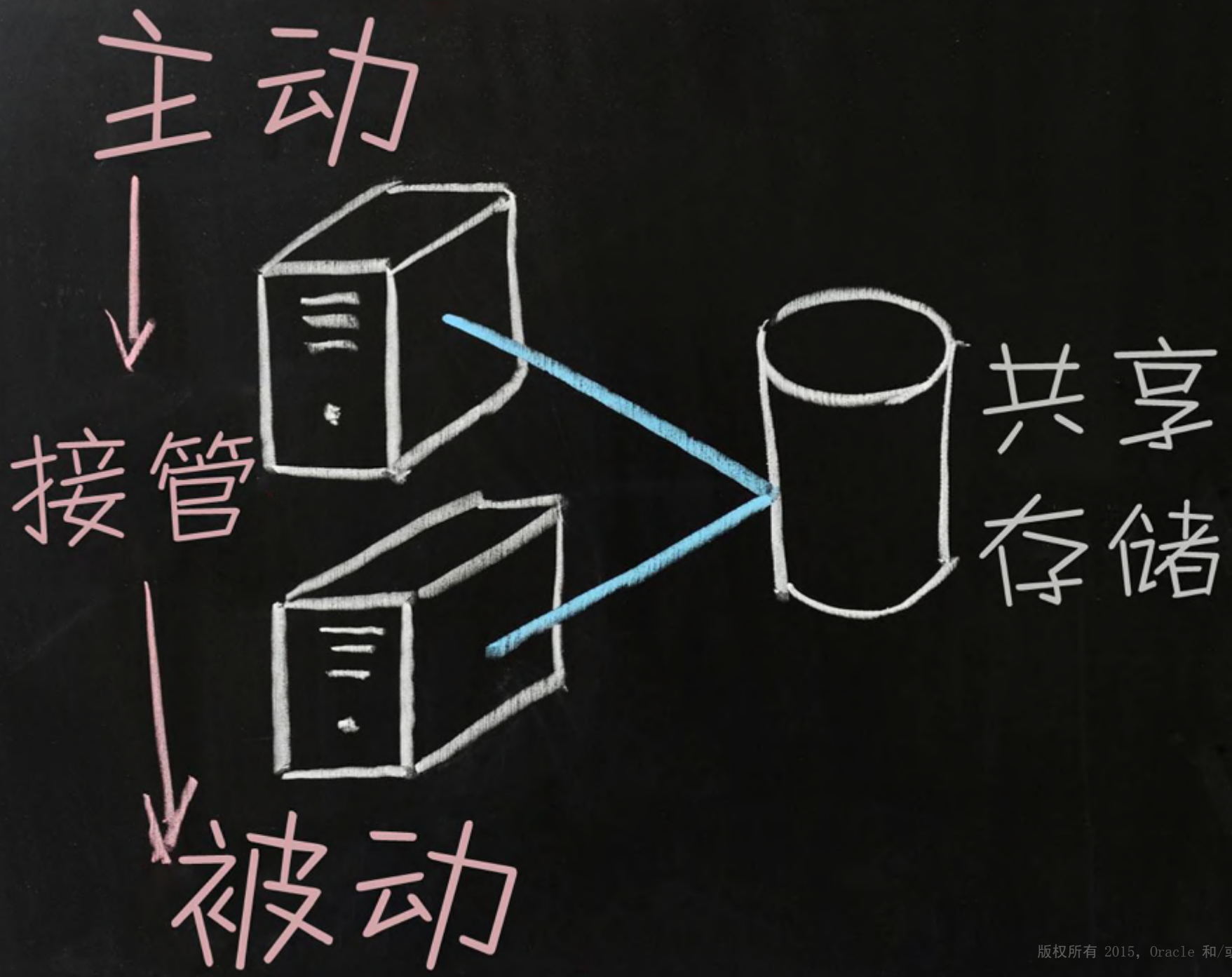
支持高可用性的各个层

并不仅仅是可靠地存储数据



- MySQL Replication
- **Shared Storage**
- Group Replication
- MySQL Cluster

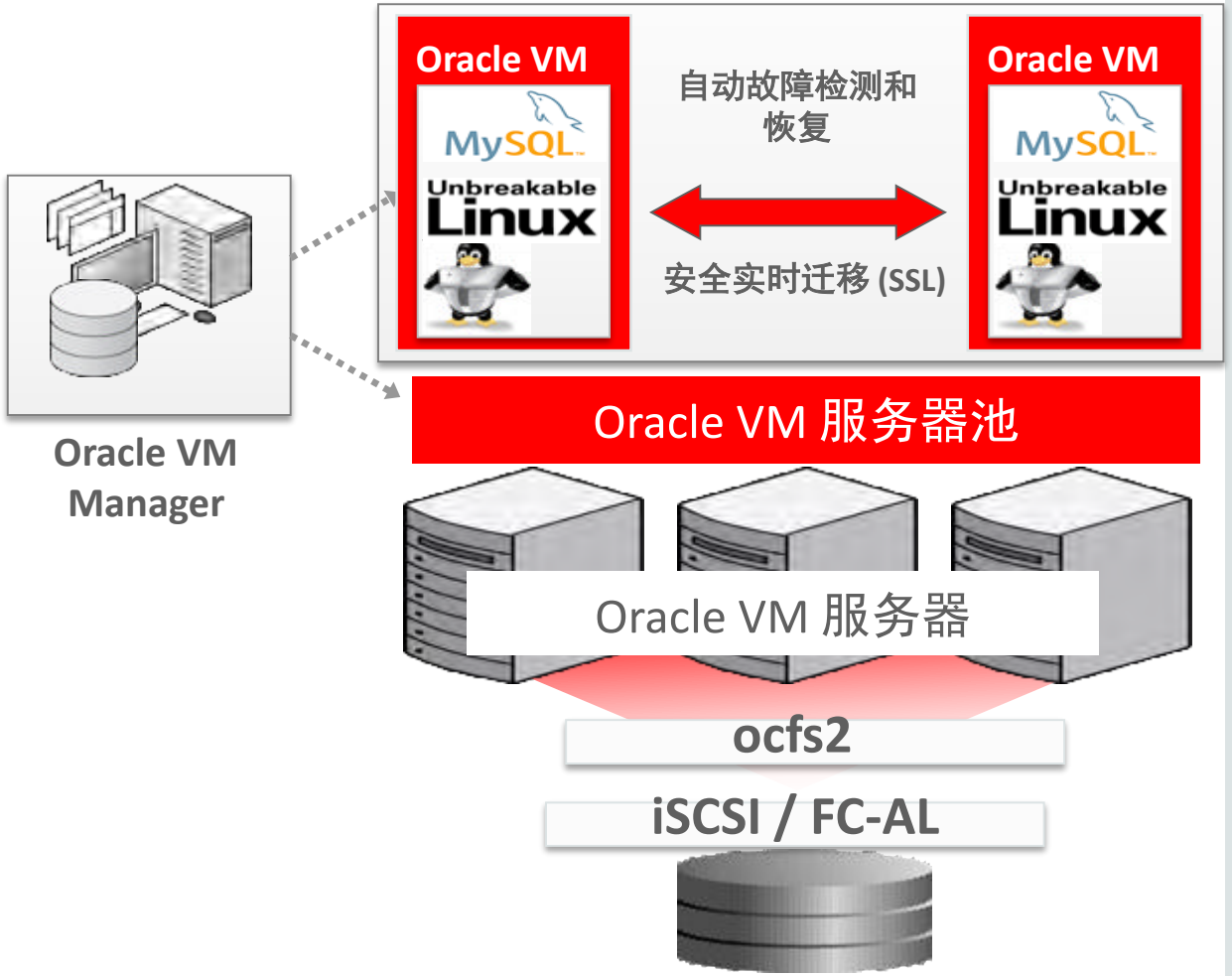
版权所有 2015，Oracle 和/或其附属公司。保留所有权利。



版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

适用于 MySQL 的 Oracle VM 模板

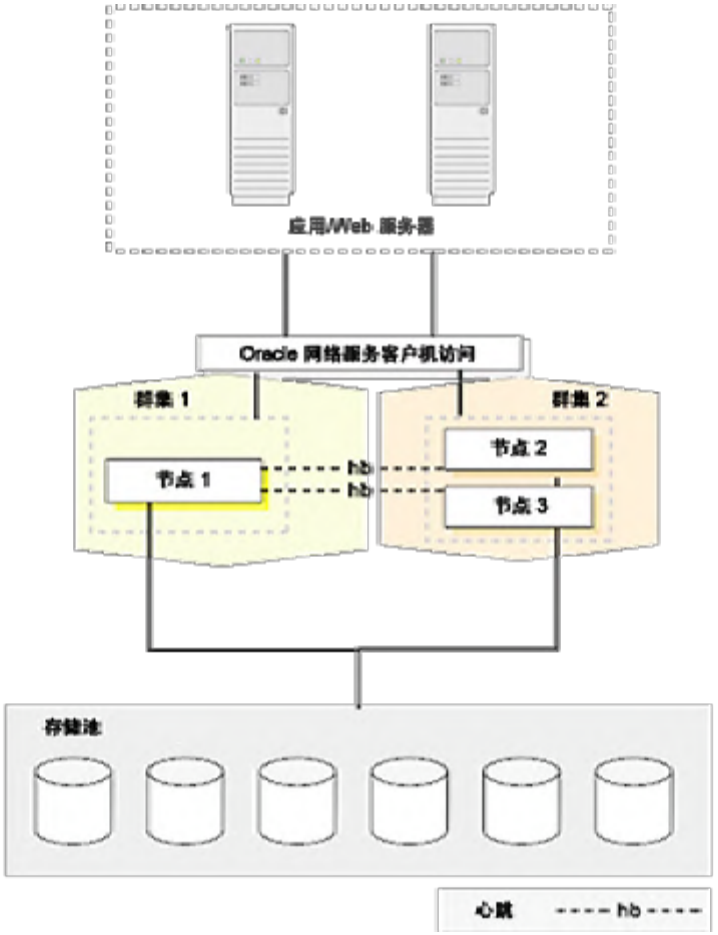
- 预安装且预配置
- 完全集成且经过 QA 测试
- 单点支持
- 通过 VM 重新启动故障实例自动恢复
- 将正在运行的实例实时迁移到新主机



版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

在 Oracle Clusterware 上部署 MySQL

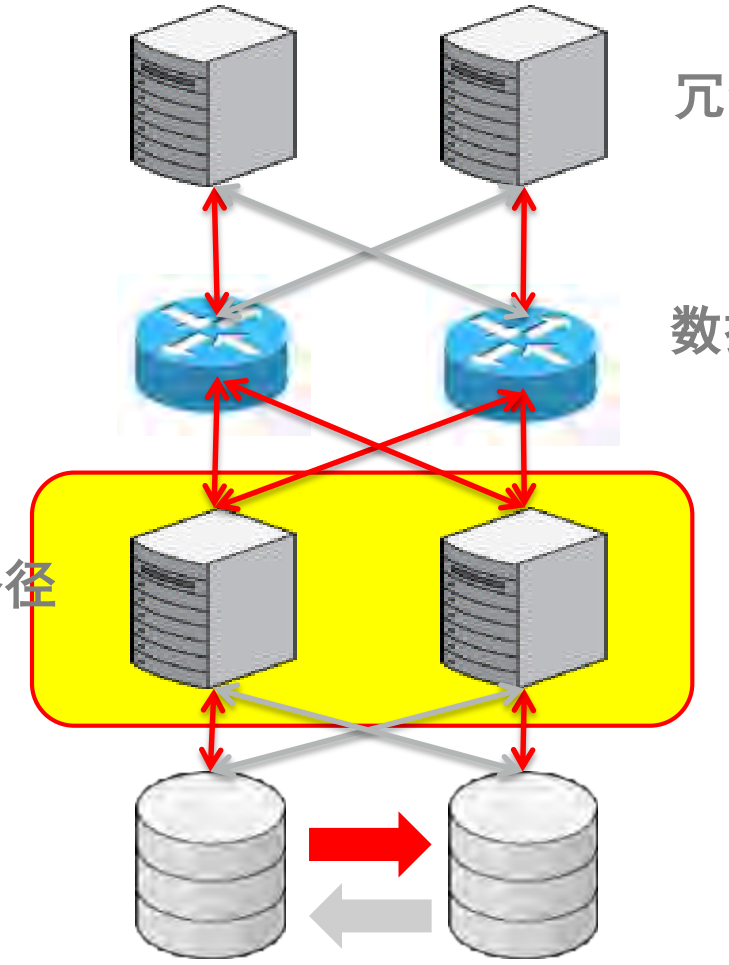
- Oracle Clusterware 将服务器场中的服务器组合为一个群集
  - 以 Oracle RAC 为核心
- Oracle Cluster 12c 包括 MySQL Server 代理
- 对 MySQL 数据库进行计划内迁移和自动故障转移
  - 对使用受管 VIP 的应用程序不可见
  - 共享存储，一次 1 个实例



版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

支持高可用性的各个层

并不仅仅是可靠地存储数据



- MySQL Replication
- Shared Storage
- **Group Replication**
- MySQL Cluster

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

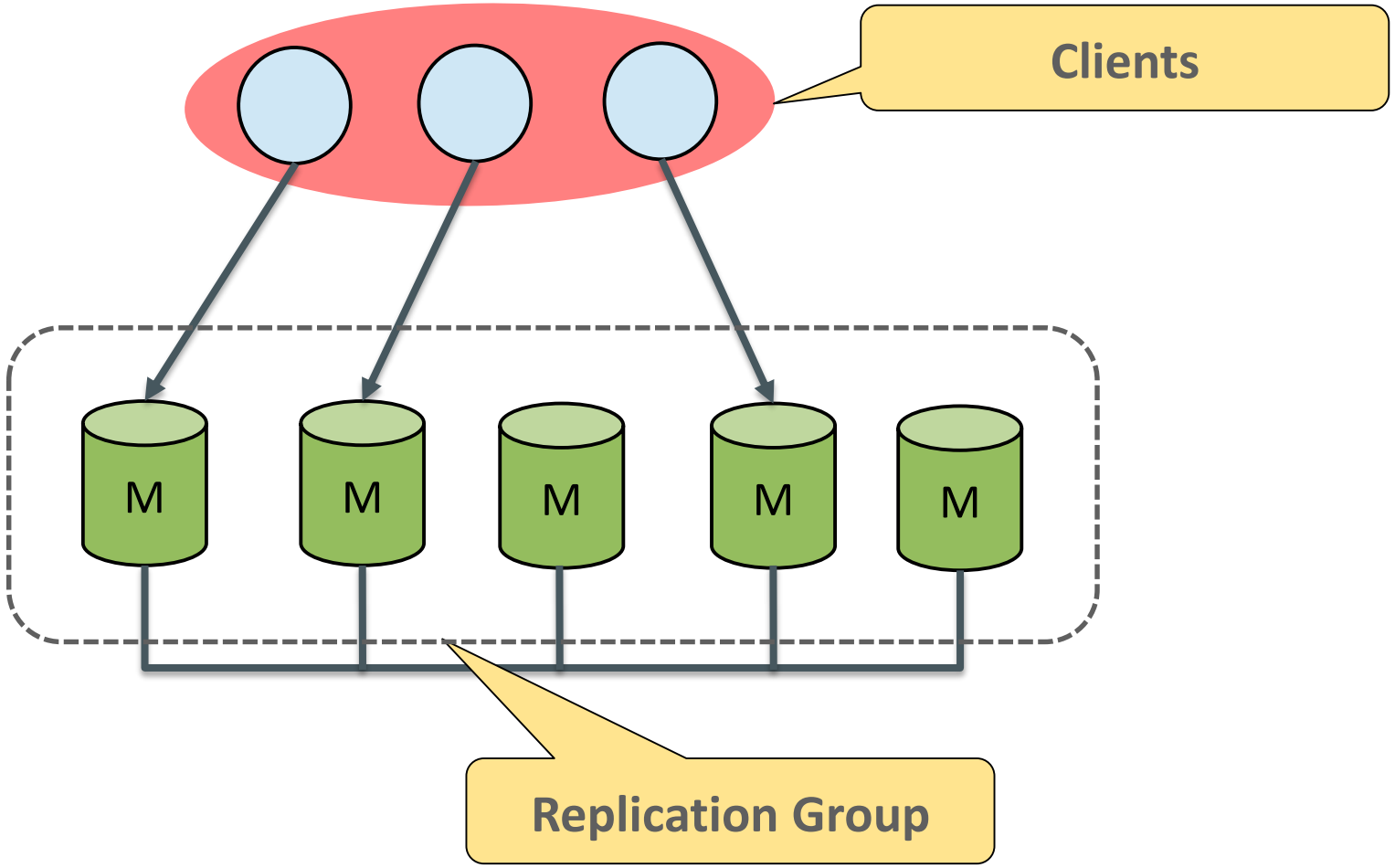
MySQL Group Replication

- MySQL Group Replication是什么?

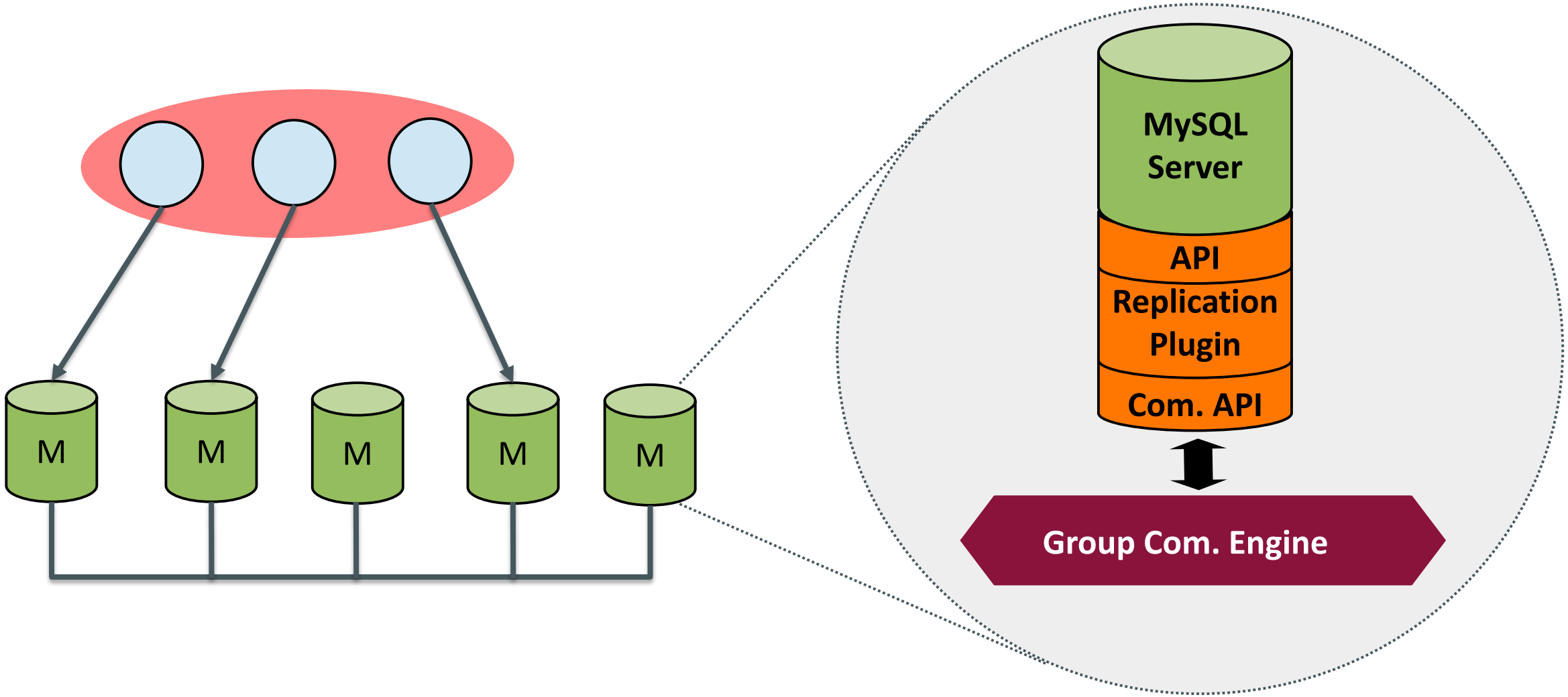
  “任何一个都可以更新 MySQL的复制插件,内建 自动分化散式的回复,冲突处理,群组成员管理 和分散式协议.”

- 能为使用者做什么?
  - 免于处理资料库的故障移转.
  - 提供容错能力.
  - 使我们能设定在群组中所有的资料库都能更新.
  - 自动化群组设定 (在宕机时,失败时,和重新连线时的处理).
  - 提供高可用复制的资料库.
  - 自动分发协调(确保不会有脑裂和讯息丢失).
  - 较少管理工作负担, 代表更大的生产力!

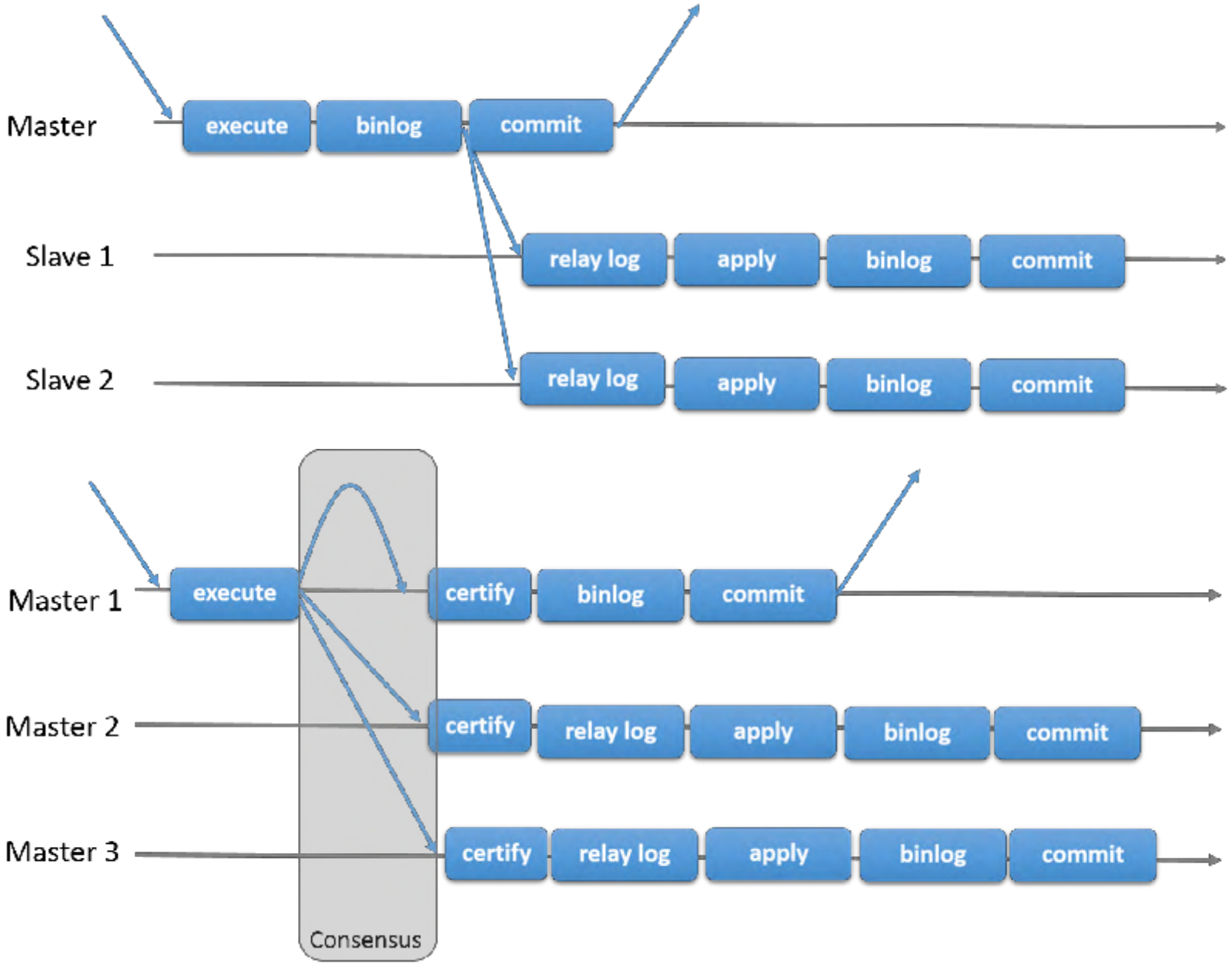
MySQL Group Replication



Major Building Blocks (1)

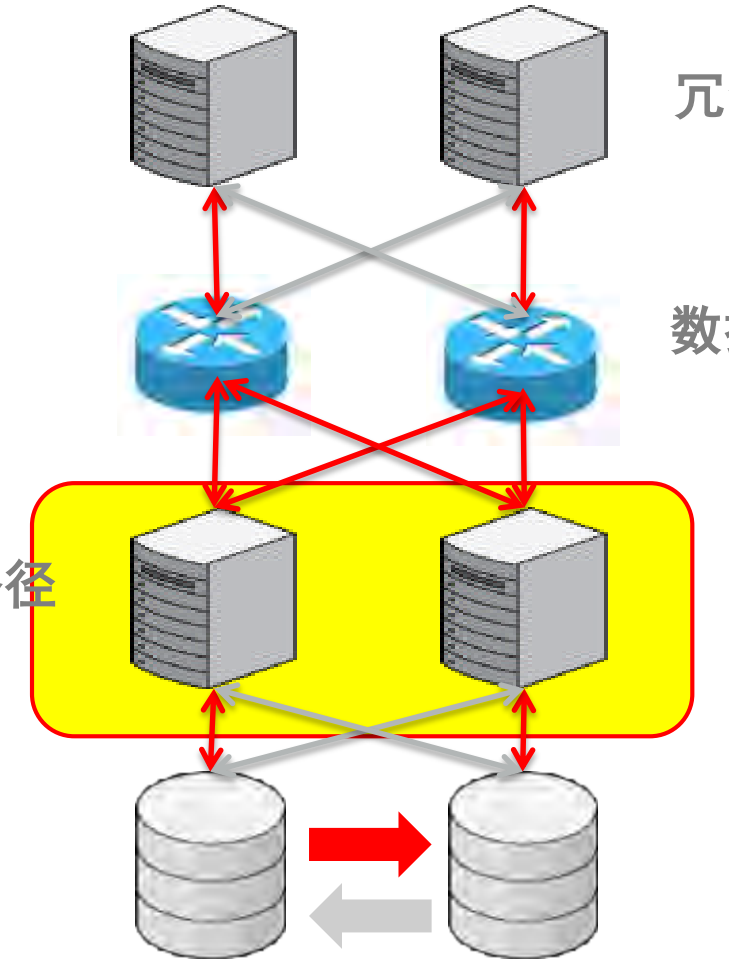


Replication vs Group Replication



支持高可用性的各个层

并不仅仅是可靠地存储数据



- MySQL Replication
- Shared Storage
- Group Replication
- **MySQL Cluster**

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

MySQL Cluster 概述

实时	• 具有持久性的内存优化表 • 延迟时间短且可预测，访问时间控制在一定范围内
高伸缩性，读取 + 写入	• 自动分片、多主 • 符合 ACID，OLTP + 实时分析
99.999% 可用性	• 无共享，无单点故障 • 自我修复 + 联机运行
SQL + NoSQL	• 关键字/值 + 复杂的关系查询 • SQL + Memcached + JavaScript + Java + HTTP/REST & C++
总体拥有成本低	• 开源 + 商业版本 • 通用硬件 + 管理、监视工具

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

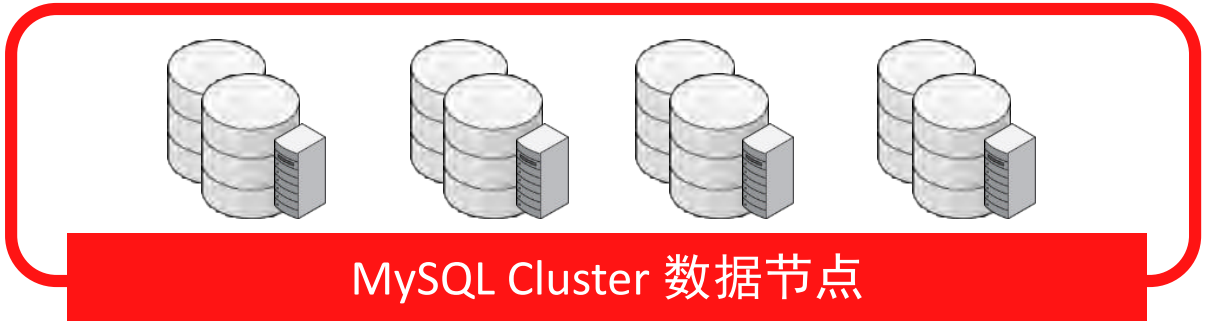
MySQL Cluster 体系结构

客户机

应用程序层

Cluster
Mgr

Cluster
Mgr

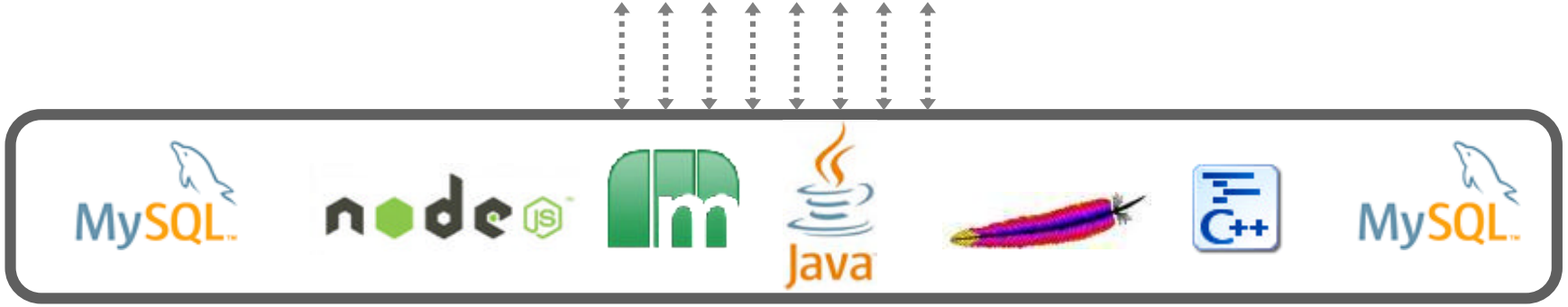


数据层

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

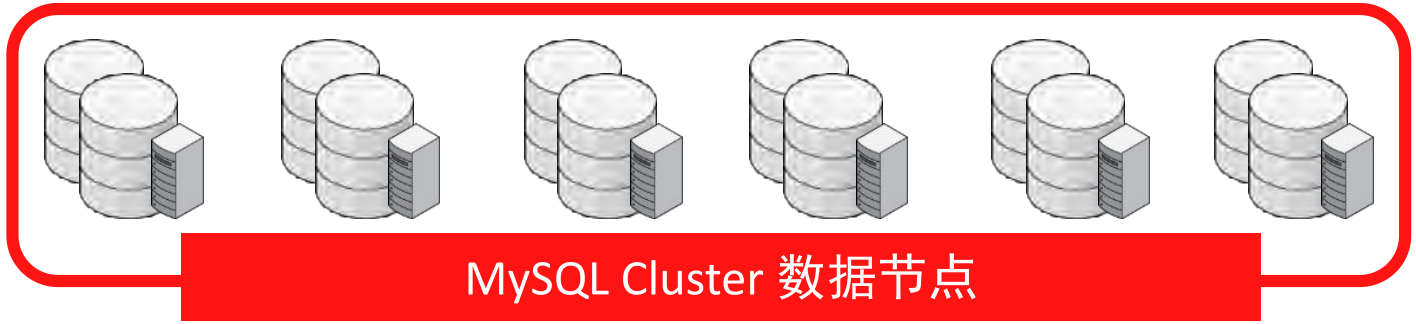
MySQL Cluster 扩展

客户机



应用程序层

Cluster
Mgr



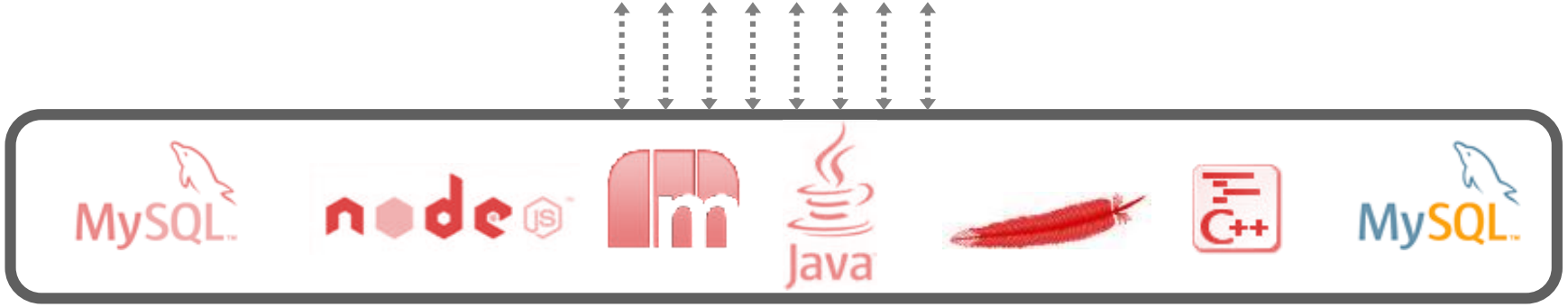
Cluster
Mgr

数据层

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

MySQL Cluster 高可用性

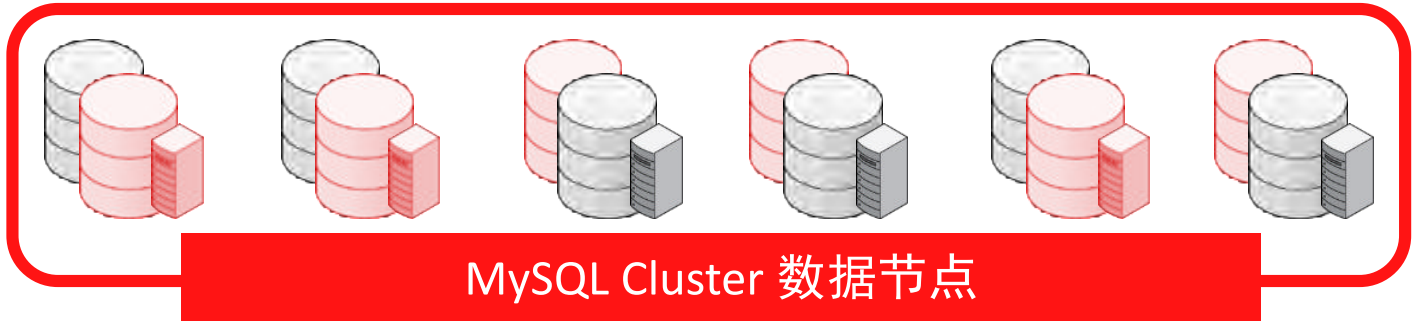
客户机



应用程序层

Cluster
Mgr

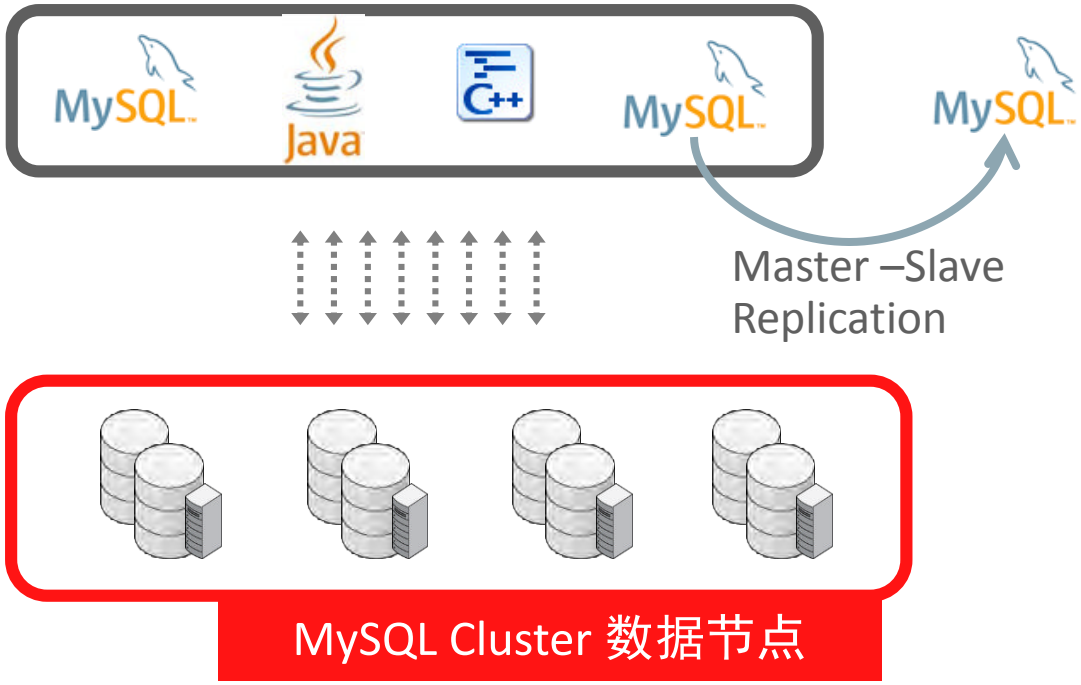
Cluster
Mgr



数据层

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

主动-主动异地复制



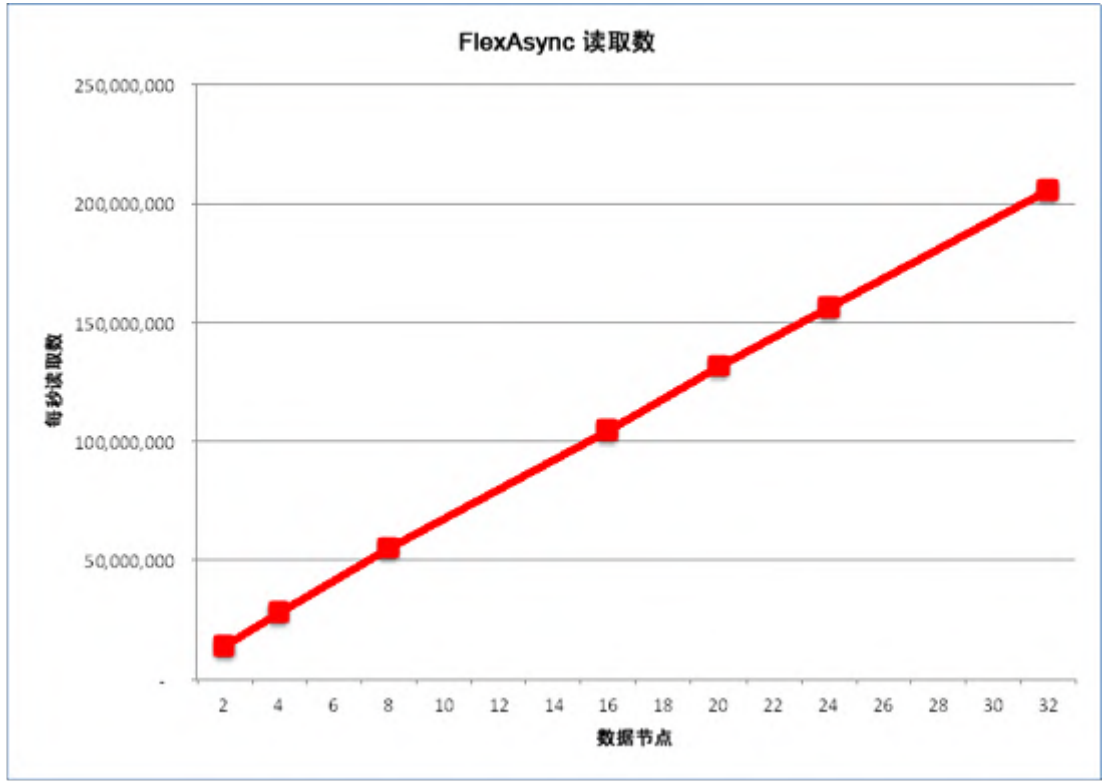
- 在 MySQL Cluster 之间异步复制
- 主动-主动
  - 随处更新
  - 冲突检测
    - 通过异常表通知应用程序
    - 可以选择自动解决冲突
  - 自动解决冲突
    - 冲突的事务和独立的事务会被回滚
- 不会对应用程序方案进行更改

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

MySQL Cluster 7.4 NoSQL 性能

2 亿次 NoSQL 读取/秒

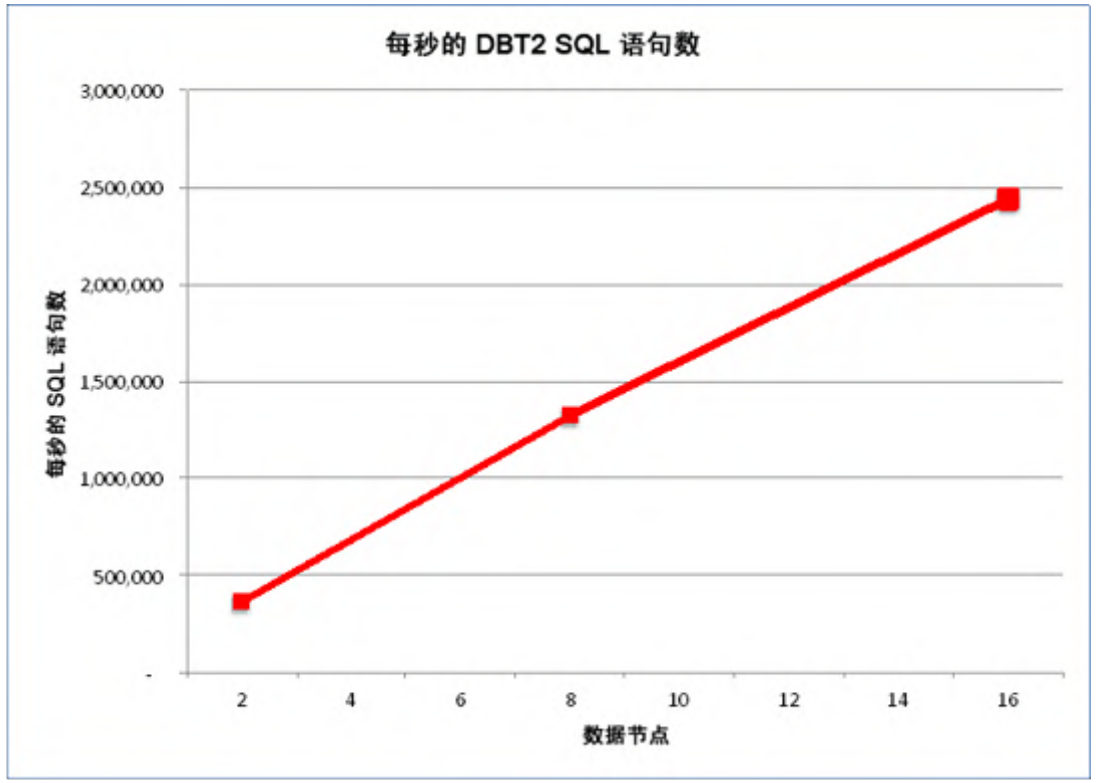
- 内存优化表
  - 持久
  - 可与基于磁盘的表混合使用
- 大量并发 OLTP
- 通过分布式联接支持分析
- 通过并行表扫描支持非索引搜索
- MySQL Cluster 7.4 DBT2 FlexAsych
  - 2 亿次 NoSQL 读取/秒（32 个节点）



版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

MySQL Cluster 7.4 SQL 性能
250 万条 SQL 语句/秒

- 内存优化表
  - 持久
  - 可与基于磁盘的表混合使用
- 大量并发 OLTP
- 通过分布式联接支持分析
- 通过并行表扫描支持非索引搜索
- MySQL Cluster 7.4 DBT2 BM
  - 250 万条 SQL 语句/秒



版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

何时考虑采用 MySQL Cluster

- 可伸缩性需求
  - 利用分片改善写入性能？
- 延迟要求
  - 每毫秒的成本？
- 正常运行时间要求
  - 每分钟的停机成本？
  - 故障与维护？
- 应用程序敏捷性
  - 开发者语言和框架？
  - SQL 还是 NoSQL？

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

增强开发运维人员的敏捷性，减少停机时间

MySQL Cluster Manager

自动化管理

- 启动/停止节点或整个群集
- 联机扩展
- 联机重新配置
- 联机升级
- 联机备份和恢复
- 导入正在运行的群集

自我修复

- 节点监视
- 自动恢复扩展到 SQL + 管理节点

高可用性操作

- 群集范围的配置一致性
- 永久性配置
- 高可用性代理

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

软件升级 — 举例说明 MCM 益处

不使用 MySQL Cluster Manager

- 群集状态初步检查需要 1 条命令
- 每台服务器 8 条 ssh 命令
- 每个进程 8 条停止命令
- 对配置文件执行 4 条 scp（2 条 mgmd，2 条 mysqld）
- 每个进程 8 条启动命令
- 检查启动和重新加入的进程需要 8 条命令
- 进程完成验证需要 8 条命令
- 验证整个群集完成恢复需要 1 条命令
- 这还不包括手动编辑每个配置文件
- 总计：46 条命令
  - 2.5 小时，有人参与

使用 MySQL Cluster Manager

MySQL Cluster Manager

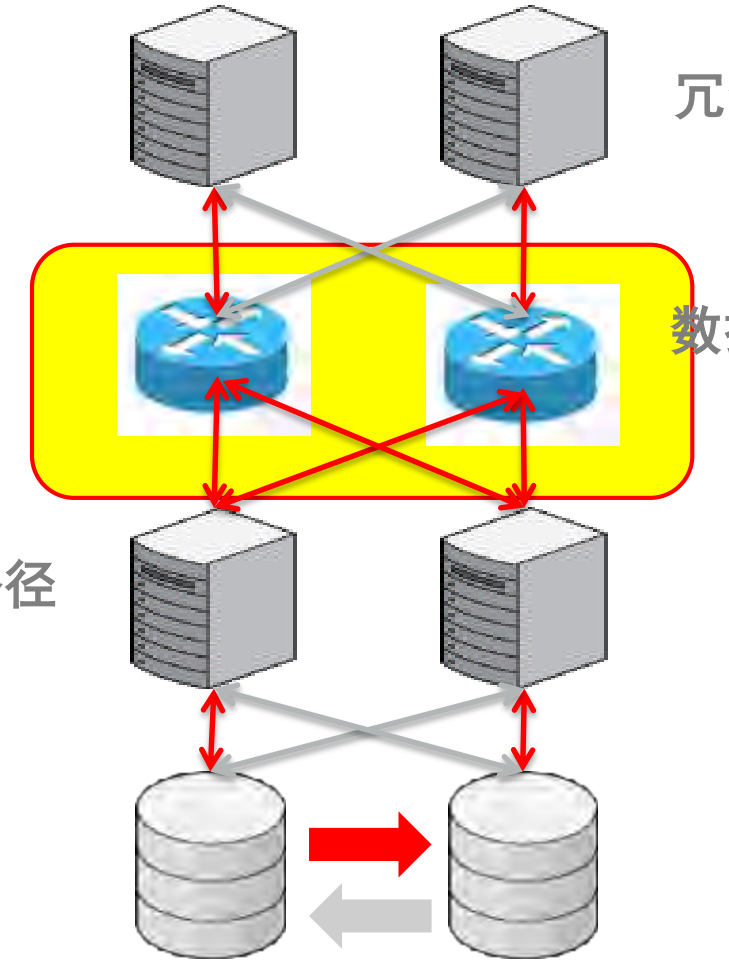
```
mcm> upgrade cluster
     --package=7.5 mycluster;
```

- 总计：1 条命令 —
  - 无人参与

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

支持高可用性的各个层

并不仅仅是可靠地存储数据

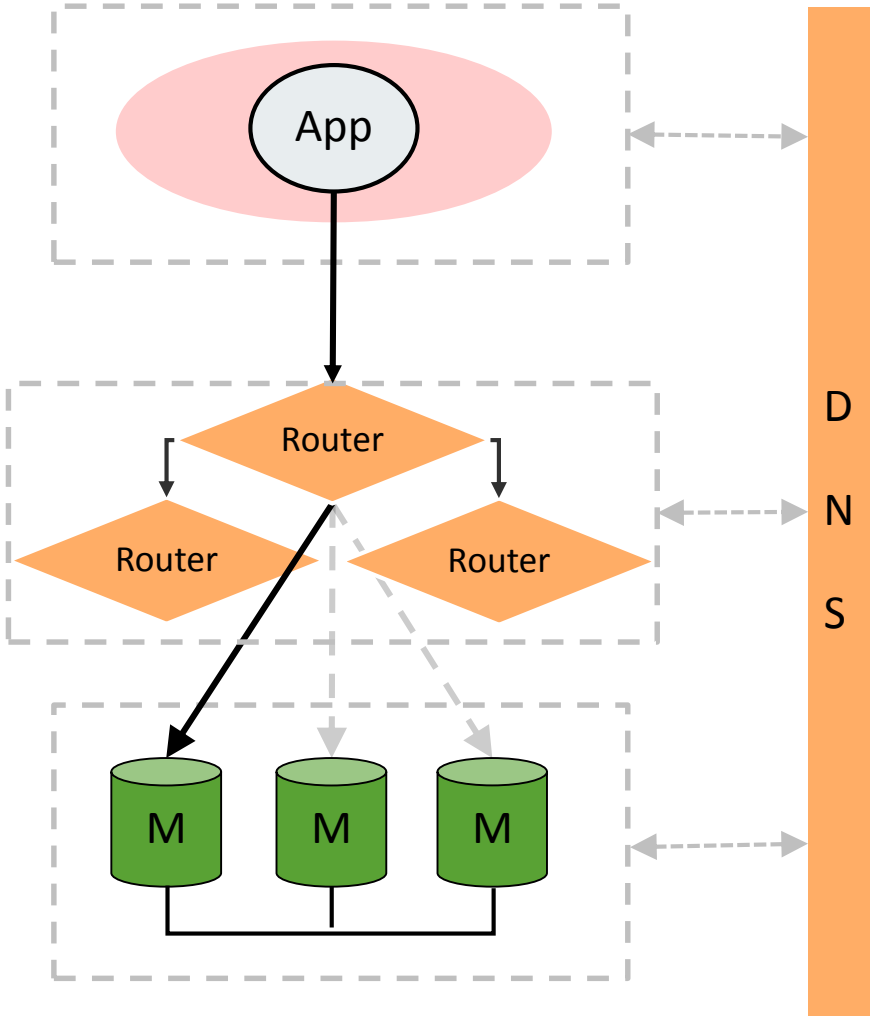


- MySQL Router
- Connector/J

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

MySQL Router HA

- Single point of failure.
- Make it redundant and a replicated state machine (using MySQL GCS).
  - Application connects using a floating host name.
  - Name server gets updated whenever the primary router fails and a new one is elected.
  - Application seamlessly fails over to the new primary router.



Oracle Confidential – Restricted

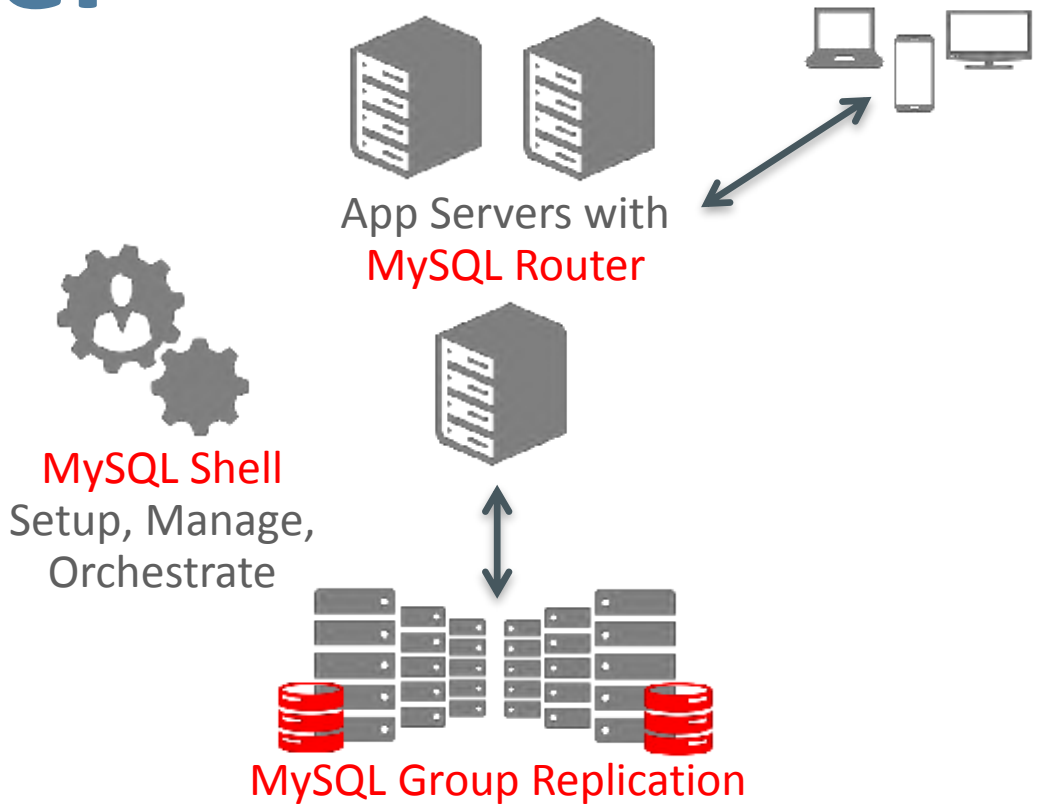
MySQL Router Configuration

```
[logger]
level = INFO

[routing:read_only]
bind_address = 127.0.0.1
bind_port = 3308
destinations = 192.168.1.141:3306, 192.168.1.42:3306, 192.168.1.43:3306
mode = read-only

[routing:read_write]
bind_address = localhost
bind_port = 3306
destinations = 192.168.1.41:3306, 192.168.1.42:3306
mode = read-write
```

ORACLE

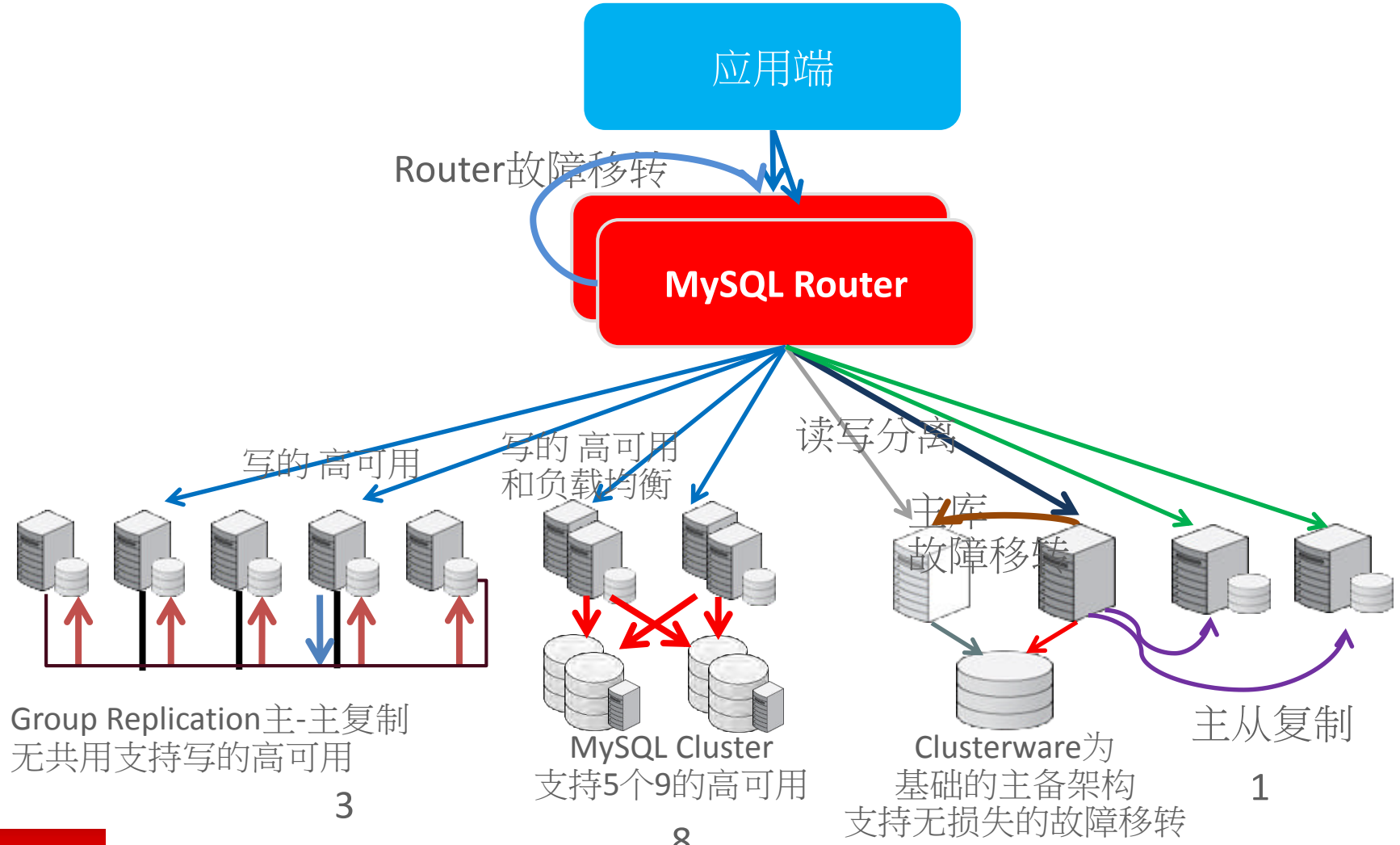
MySQL InnoDB Cluster

"High Availability becomes a core first class feature of MySQL!"



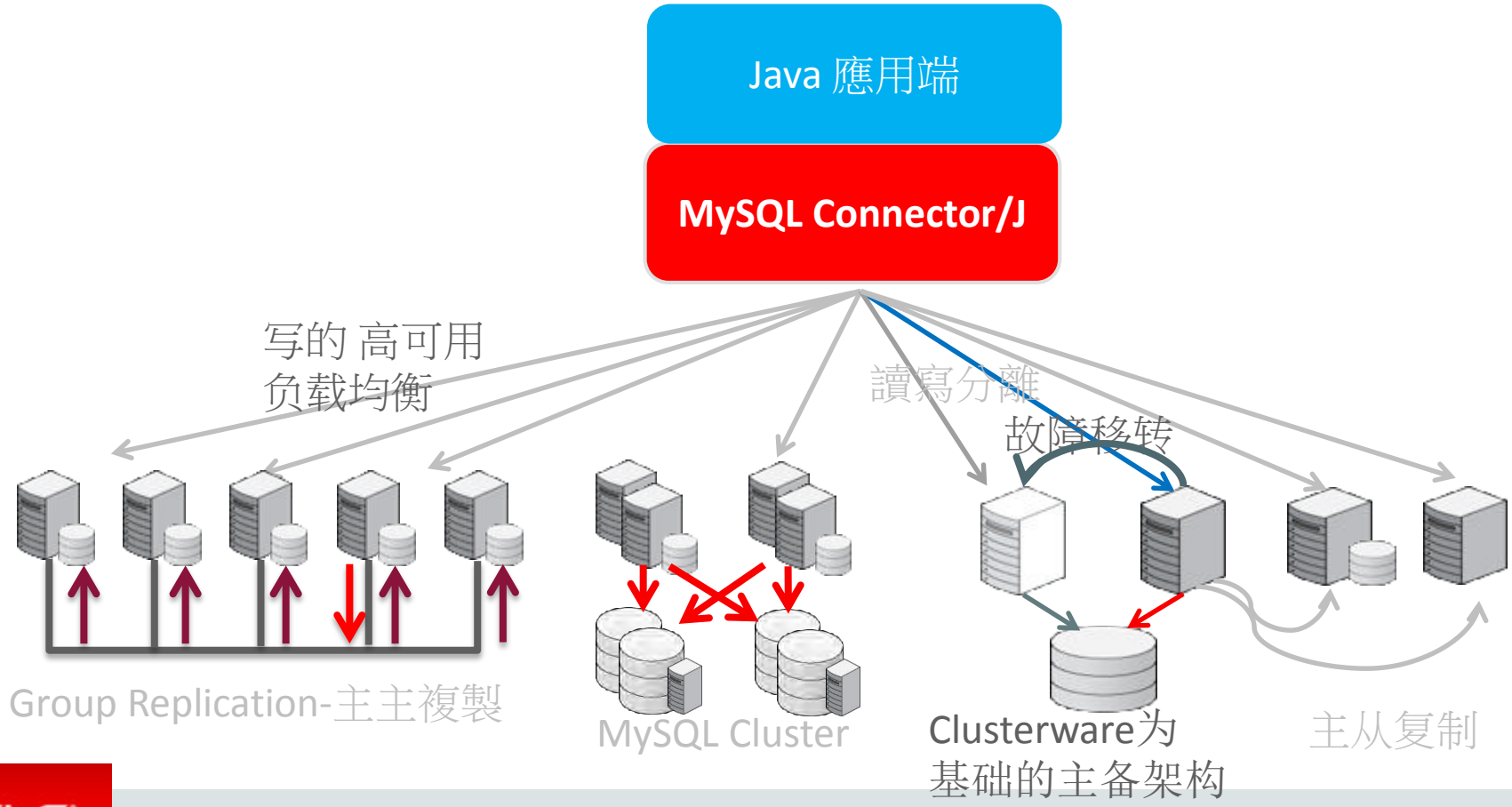
MySQL Router应用能透明的和数据库连结

-支持负载均衡,高可用读写分离和数据分片



Connector/J Failover

-支持负载均衡,高可用读写分离和数据分片



ORACLE

故障移转

- JDBC URL format，第一个主机为master，第二个以后为backup:

```
import com.mysql.jdbc.ReplicationDriver;

Connection conn =
  DriverManager.getConnection('jdbc:mysql://primaryhost:3306,host2:3306,host3:330
  6…?[propertyName1=propertyValue1[&propertyName2=propertyValue2]...]
```

- Connection的faileover属性
  - failOverReadOnly :为true时，faileover后设Connection.setReadOnly(false)；第二个host仍为read-only mode
  - secondsBeforeRetryMaster
  - queriesBeforeRetryMaster
  - retriesAllDown
  - autoReconnect
  - autoReconnectForPools

ORACLE

Oracle MySQL 高可用性和扩展解决方案

	MySQL 复制	Oracle VM MySQL 模板	Oracle Clusterware	Solaris Cluster	Group Replication	InnoDB Cluster	MySQL Cluster
应用程序自动故障转移	✖	✔	✔	✔	✖	✔	✔
数据层自动故障转移	✖	✔	✔	✔	✔	✔	✔
零数据丢失	MySQL 5.7	✔	✔	✔	✔	✔	✔
平台支持	全部	Linux	Linux	Solaris	全部	全部	全部
群集模式	主 + 从	主备	主备	主备	多主	多主	多主
故障转移时间	N/A	数秒以上	数秒以上	数秒以上	1 秒以上	1秒以上	< 1 秒
扩展	读取	✖	✖	✖	读取	读取	✔
跨分片操作	N/A	N/A	N/A	N/A	N/A	N/A	✔
透明路由	✖	✔	✔	✔	✔	✔	✔
无共享	✔	✖	✖	✖	✔	✔	✔
存储引擎	InnoDB+	InnoDB+	InnoDB+	InnoDB+	InnoDB+	InnoDB+	NDB
单一供应商支持	✔	✔	✔	✔	✔	✔	✔

版权所有 2015，Oracle 和/或其附属公司。保留所有权利。

关注QCon微信公众号，
获得更多干货！

Thanks!

INTERNATIONAL SOFTWARE DEVELOPMENT CONFERENCE

主办方 Geekbang. 极客邦科技 InfoQ

ORACLE®