云原生的 MySQL 托管服务架构及读写分离的优化

宋青见
C+E CCIC
Principle PM

Microsoft

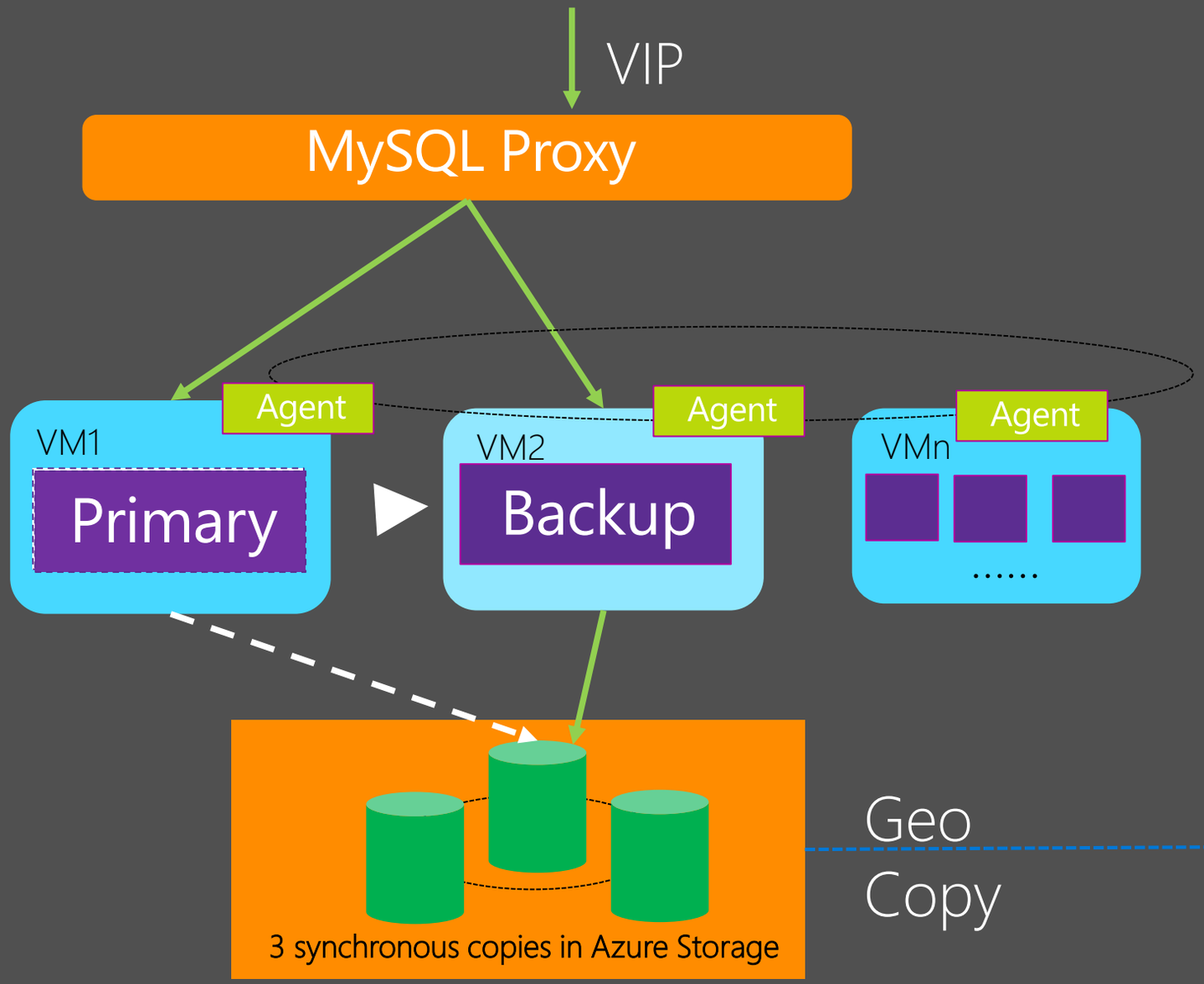
议题

- 云原生的 Azure RDS for MySQL 托管服务架构
- 读写分离的优化
- 微服务架构Service Fabric的相关介绍

Microsoft

云原生的托管服务架构 – DevOps（开发工程师运维）

云原生的 MySQl PaaS服务： 高可用高可靠



基于Azure存储提供数据的高可用性和高可靠性

本地3份同步数据拷贝
异地3份异步数据拷贝

数据库服务的高可用

设计和运维以99.9%高可用为标准
Enable-Secondary 启用备用库
达到99.99%

支持异地灾备恢复

下一步的架构：一体化的数据库运维平台 已全球上线

Power BI, App Services, Data Factory, Analytics, ML, Cognitive, Bot…

SQL Data Warehouse	SQL Database	MySQL 预览	PostgreSQL 预览
Azure 关系型数据库托管运维平台	智能: 建议, 优化, 监测		
	灵活: 按需弹性放缩，资源管理		
	可信: 高可用/灾备 备份/恢复, 安全, 审计, 多租户隔离		

Azure 云主机+网络

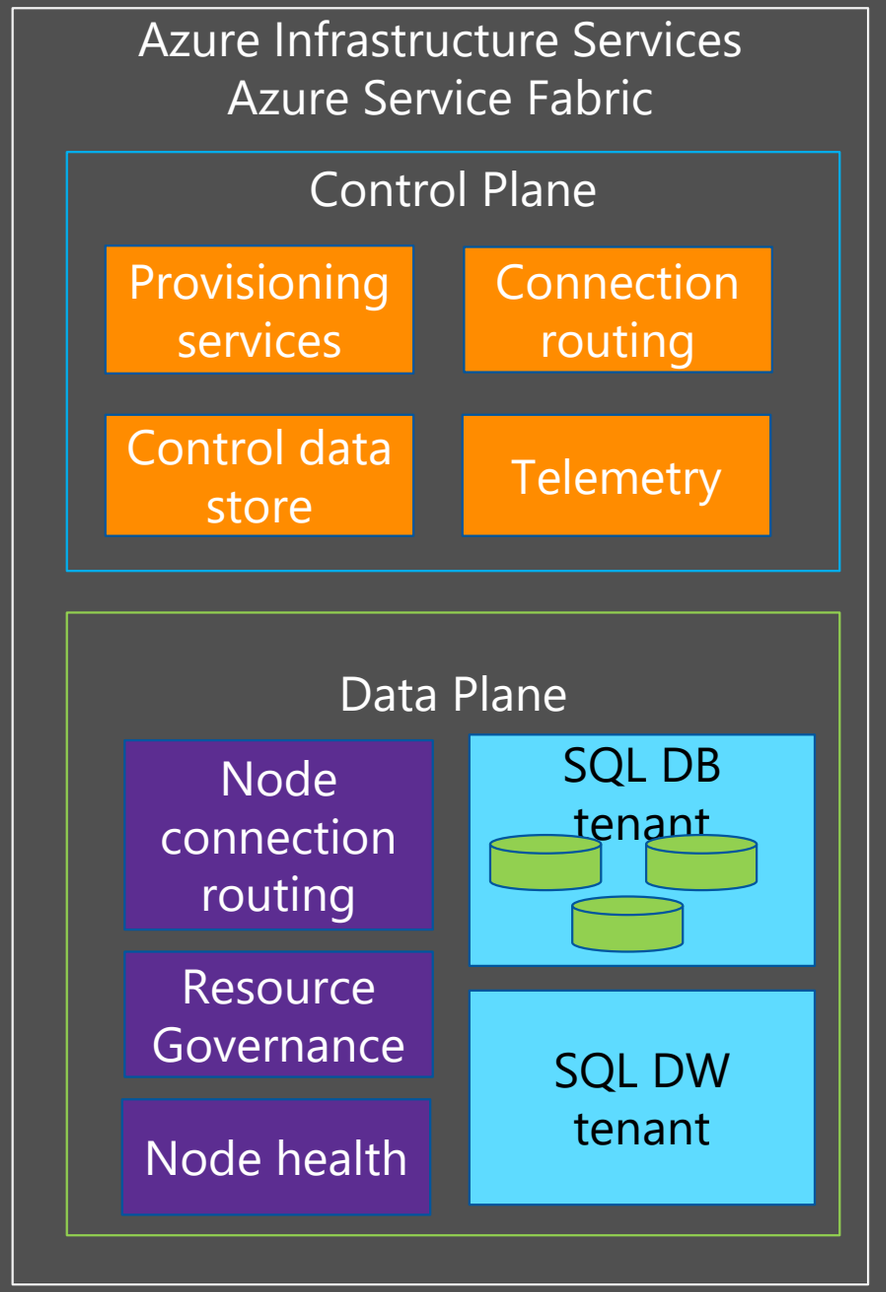
Azure 云存储

全球38个数区域据中心

智能 // 可信 // 灵活

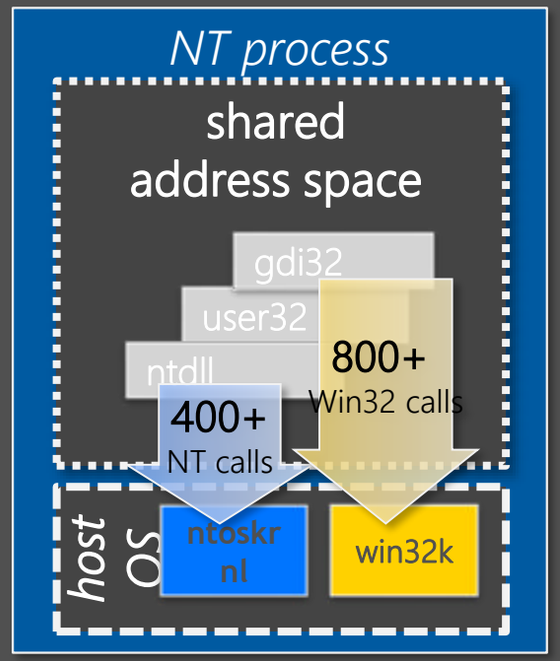
Azure Global Database Service architecture

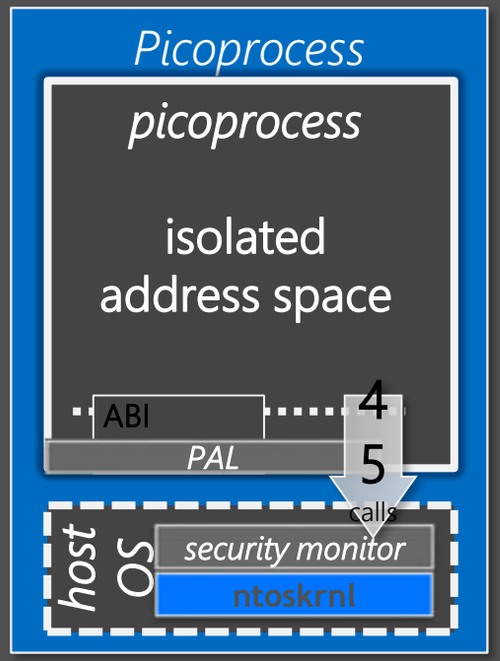
- DB Cluster is decomposed into Azure Service Fabric applications
- All applications and all DB tenants are individually deployable
- Databases are “services” managed by Azure Service Fabric



Drawbridge: A container technology to achieve isolation, security and density in the cloud

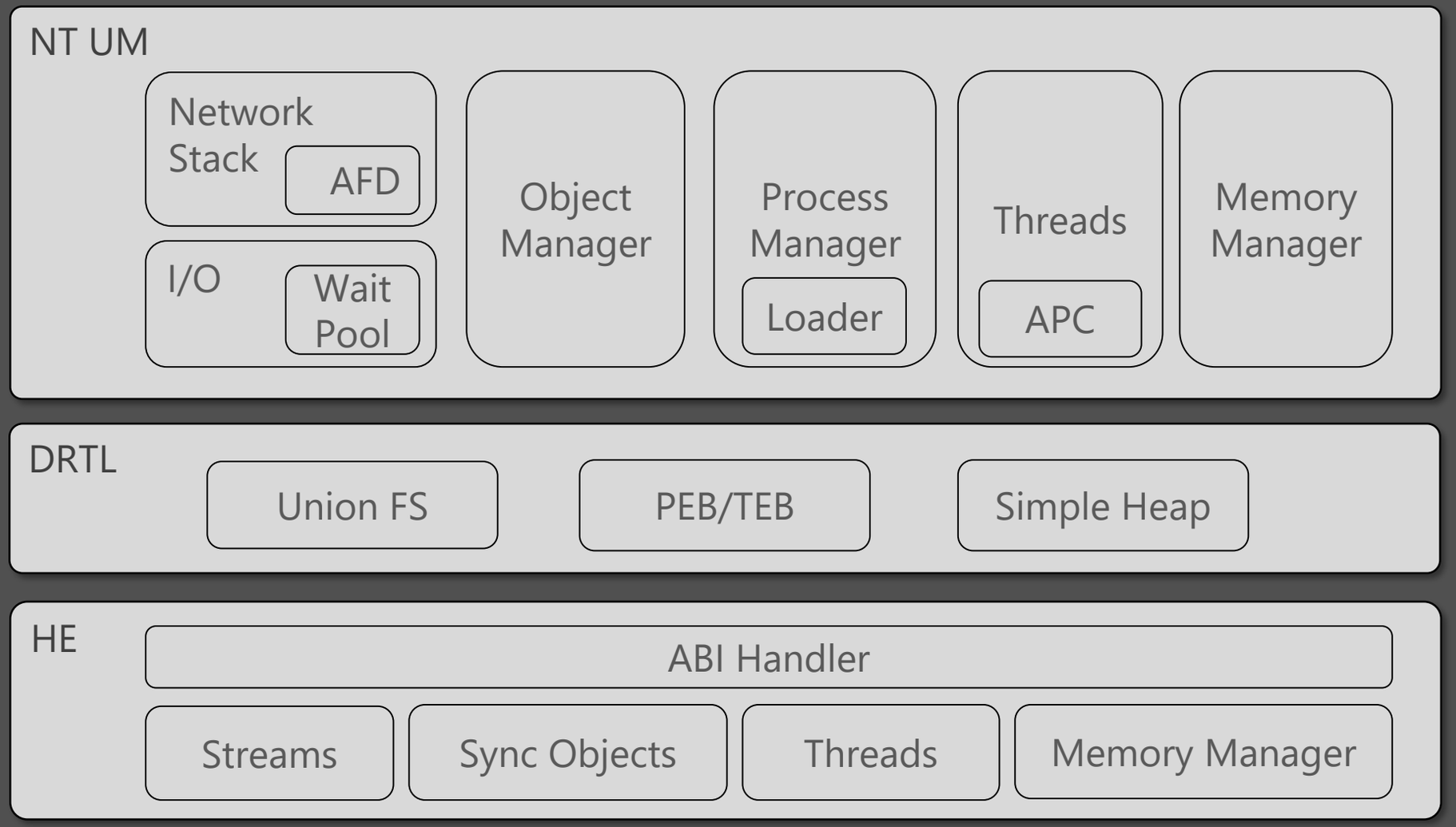
- Modified Windows Kernel to run in user mode, aka Library OS or LibOS
- Designed for running on Windows and leverages Pico-process feature
- Pico-process is a NT process with empty address space



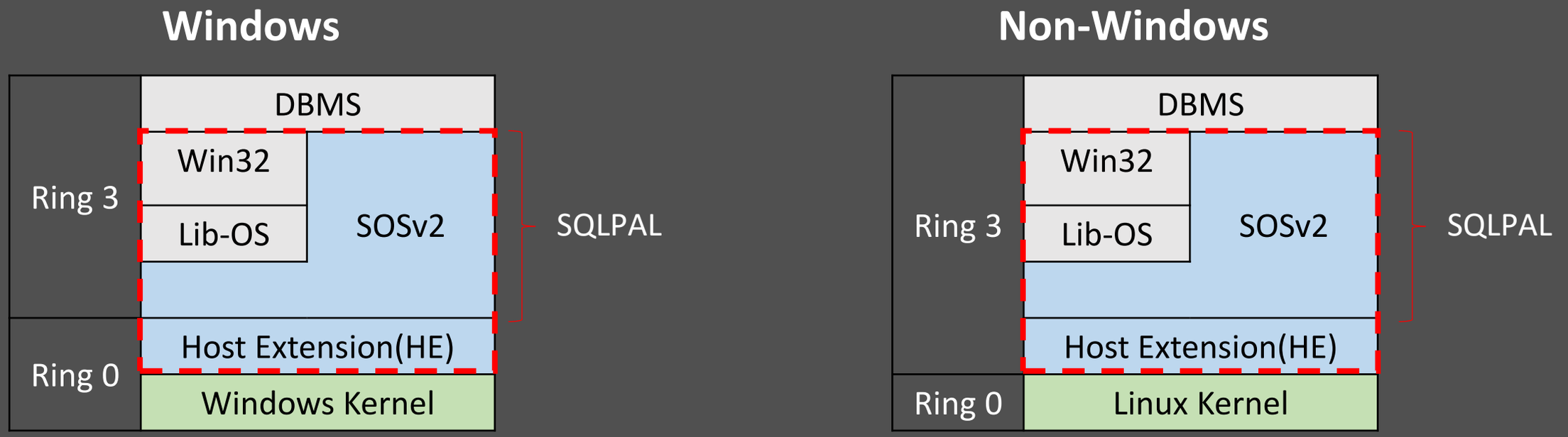


- All 1200+ system calls blocked from user-mode (NTOS and win32k)
- Enforced by 35-line change to KiSystemServiceHandler
- No perf impact to other processes —leverages “slow path” used by UMS
- 45 new system calls added to process (Drawbridge system calls)
- Even hard-coded traps can’t break out

LibOS: A user mode runtime library exposing semantics of Windows kernel

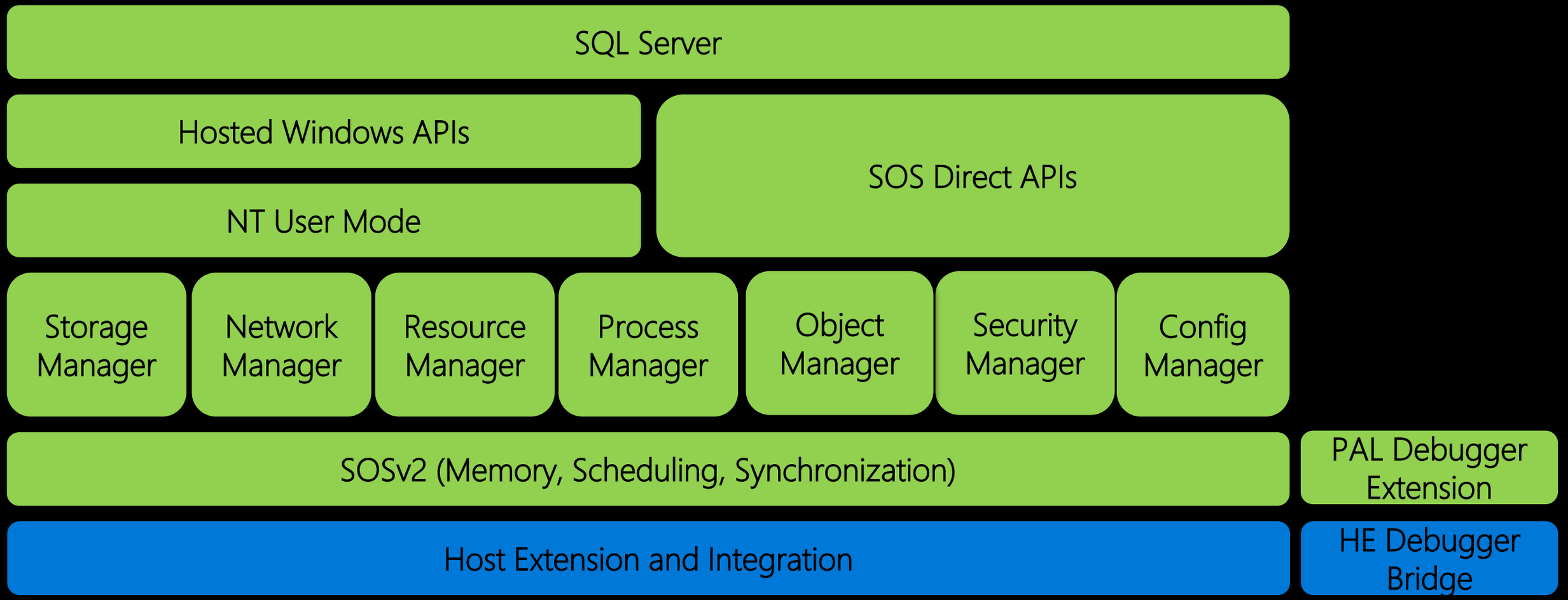


SQL Platform Abstraction Layer (SQLPAL): Windows and Linux



- Windows Host Extension has a driver for creating the Pico process and a monitor process (user mode) that implements non-perf related ABIs. ABI calls are handled by the driver and are either handled directly (Like File IO) or are marshalled to the monitor process for handling (like File Open)
- On Linux everything is in user mode. Main difference is Ring 0 to 3 transition point. And hence no isolation

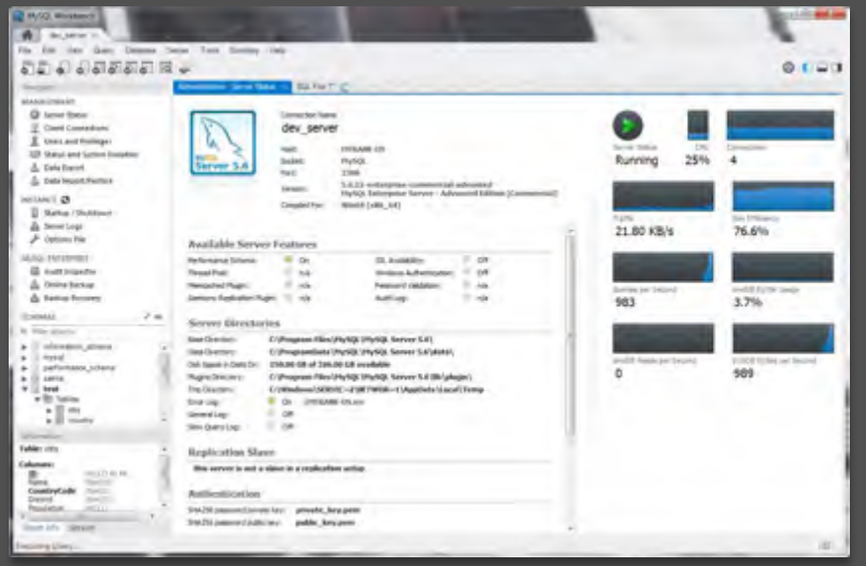
SQL PAL and SOSv2 Architecture



支持熟悉的平台和工具

后台用的DB Engine是MySQL 社区版本（Community Edition）

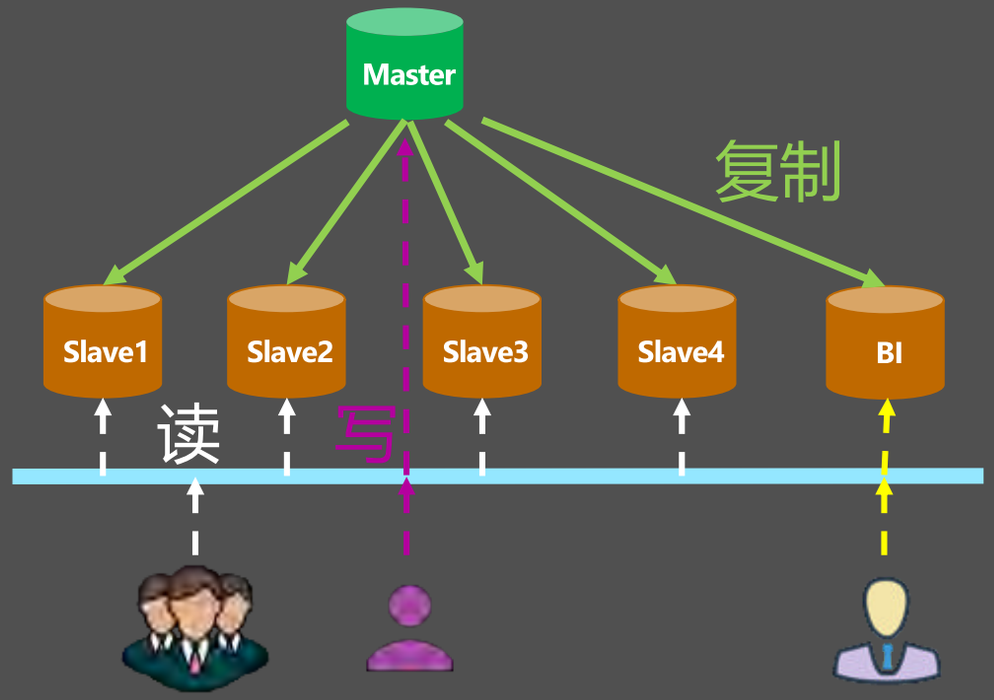
支持现有的MySQL客户端和工具 (例如phpMyAdmin, MySQL workbench, navicat 等）



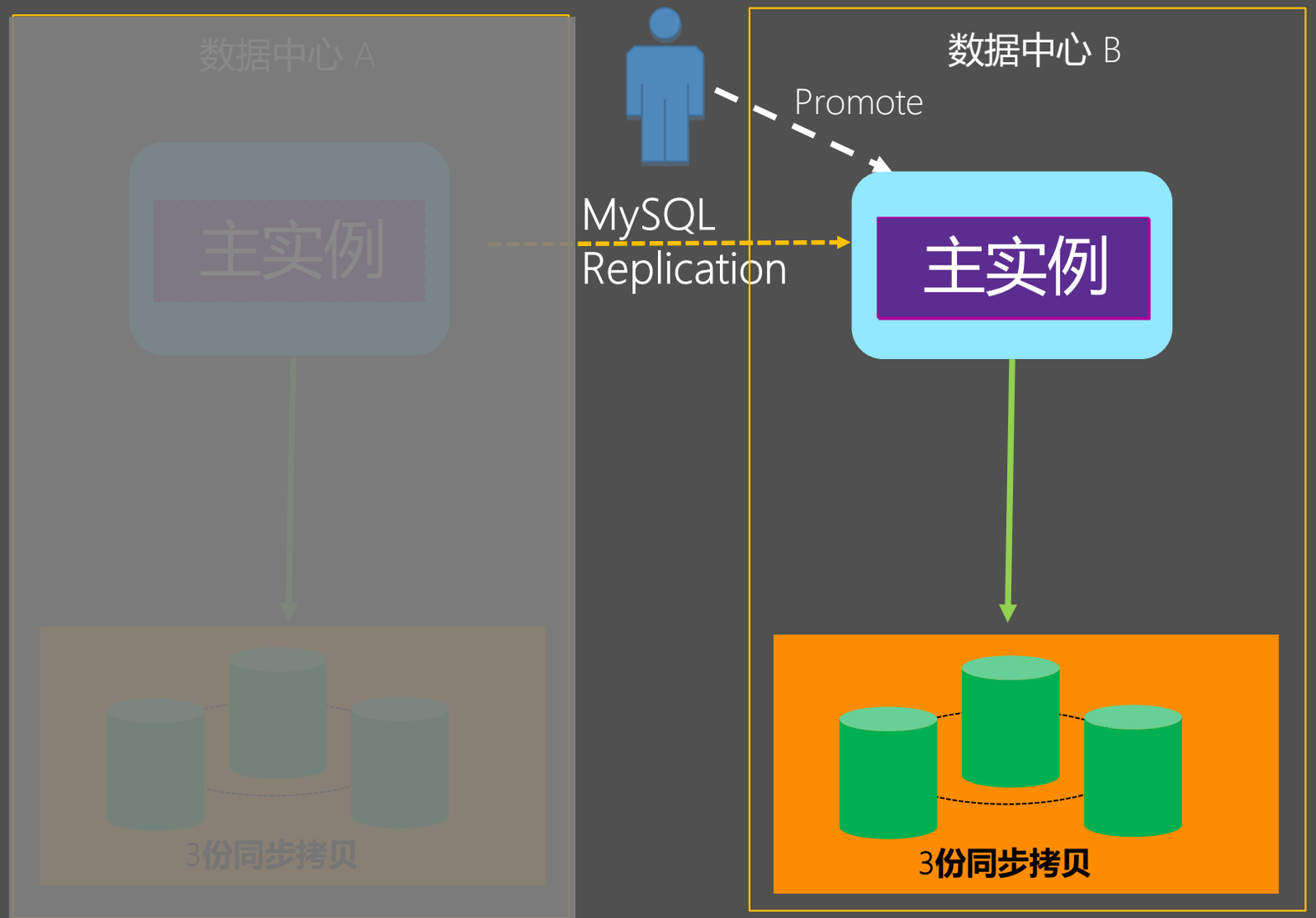
读写分离的优化

读写分离的实现细节和优化

- 关闭MySQL本身的主从复制功能
  - 主库的 Binlog 已经存储于 Azure Storage
  - 在PaaS内设置 一主多从，不需要打开MySQL本身的主从复制功能
- 通过外部独立进程，进行主库binlog的解析和从库入库操作
  - Commit 的性能损耗 > Write的性能损耗
  - Write Combine 优化



灾备恢复 – 基于异地副本（replica）



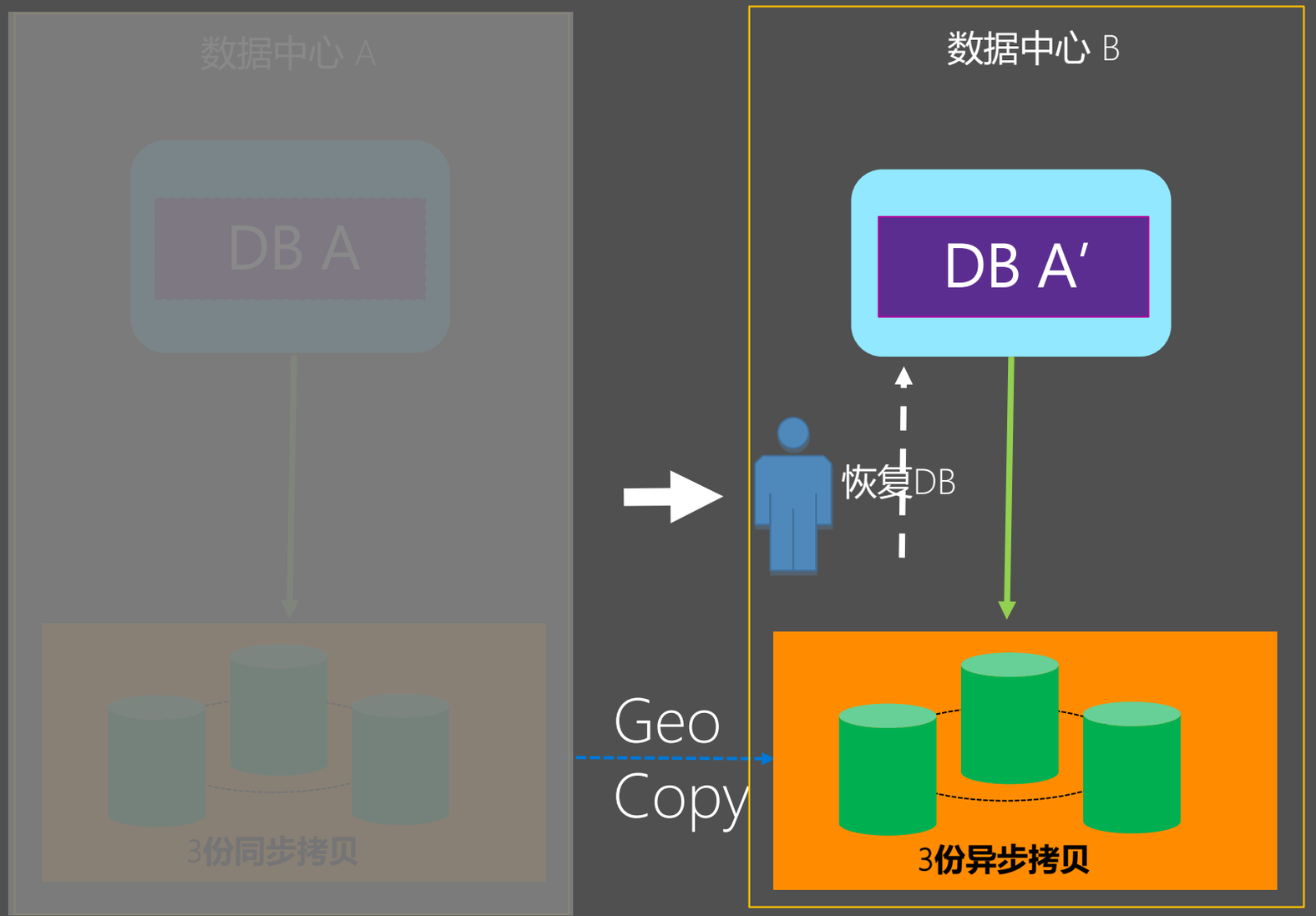
基于异地副本的恢复

灾备恢复指标
RPO < 30秒， ERT < 10秒

运维人员通过Portal或Powershell升级副本为主实例

需要支付异地副本的费用

灾备恢复 – 基于异地数据库恢复



基于异地数据拷贝的恢复

灾备恢复指标

RPO < 1小时， ERT < 3小时

运维人员通过Powershell进行恢复

所有版本具备这个功能，没有额外费用

支持混合云的数据库同步

支持标准的MySQL Slave模式

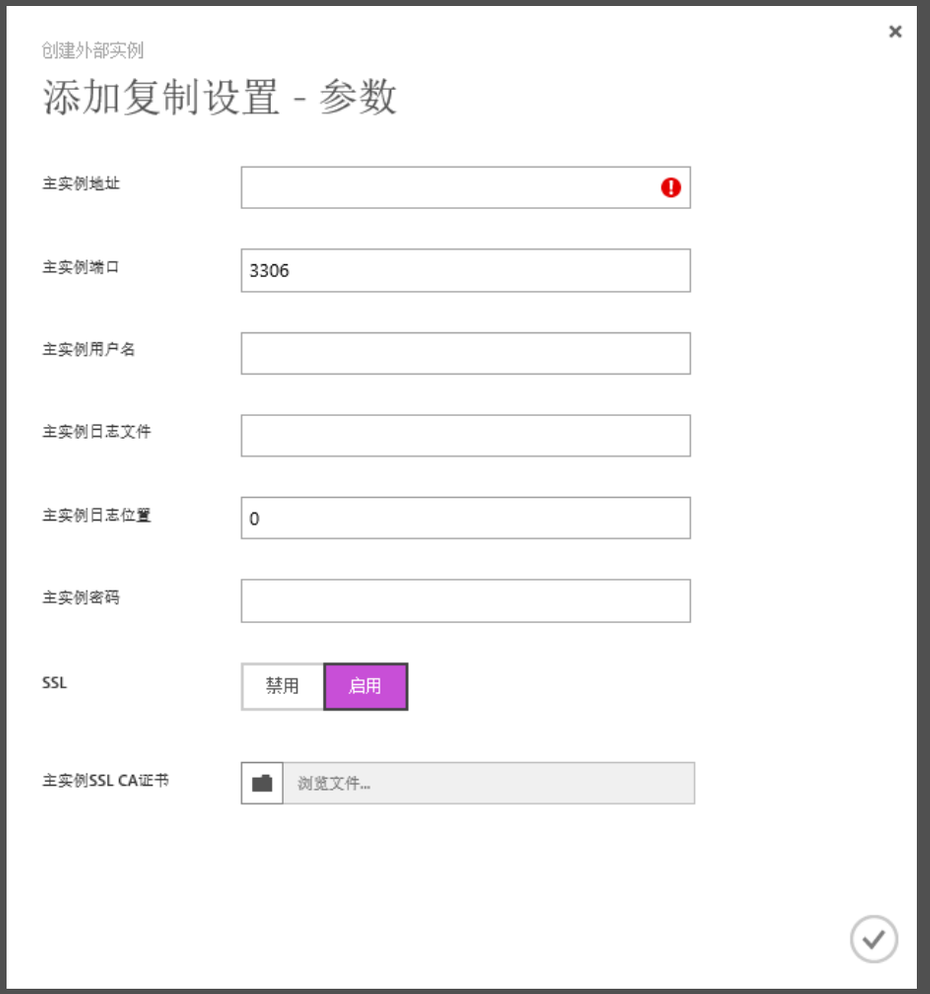
常见混合云场景

- 从本地同步数据库到Azure上以满足Azure上的应用需要
- 支持应用平滑迁移到Azure

通过管理门户配置同步和查看同步状态

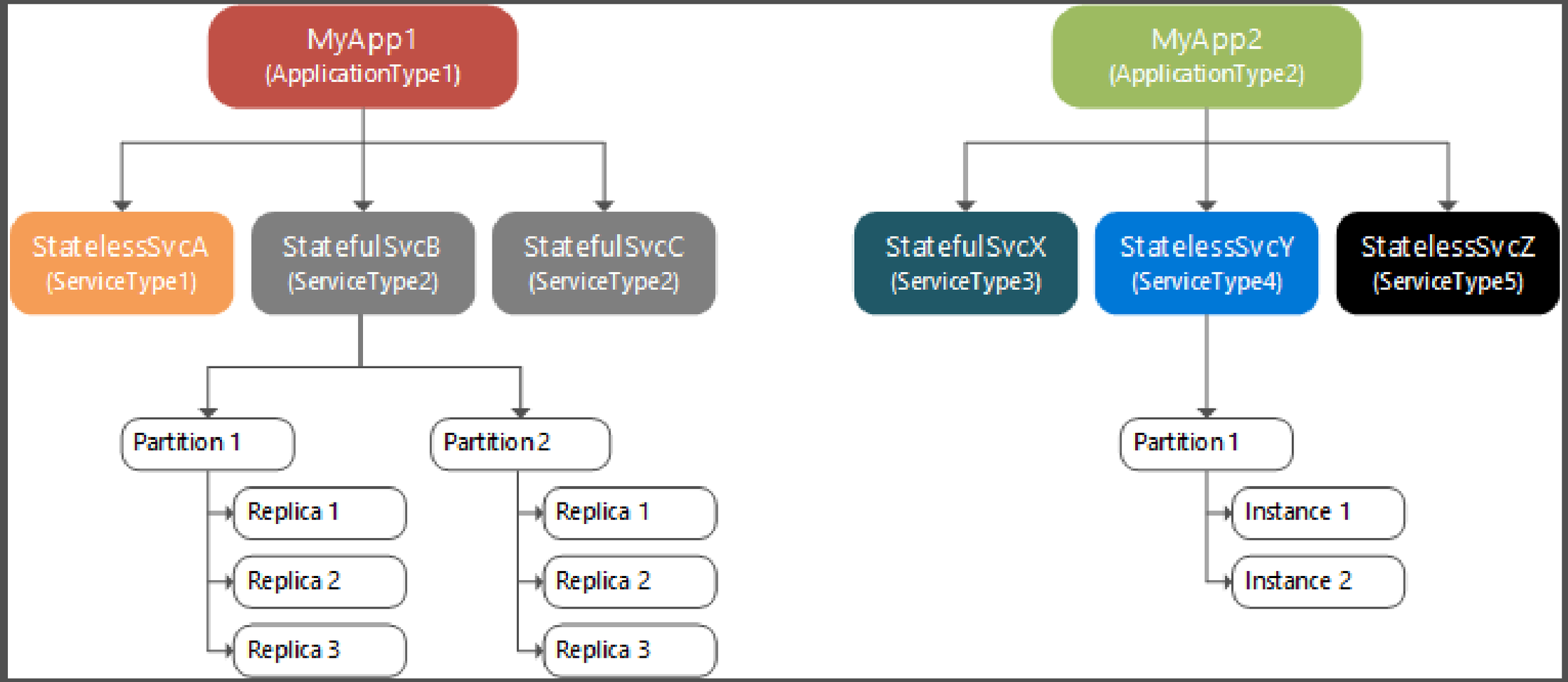
配置文档

- http://www.windowsazure.cn/documentation/articles/mysql-database-data-replication

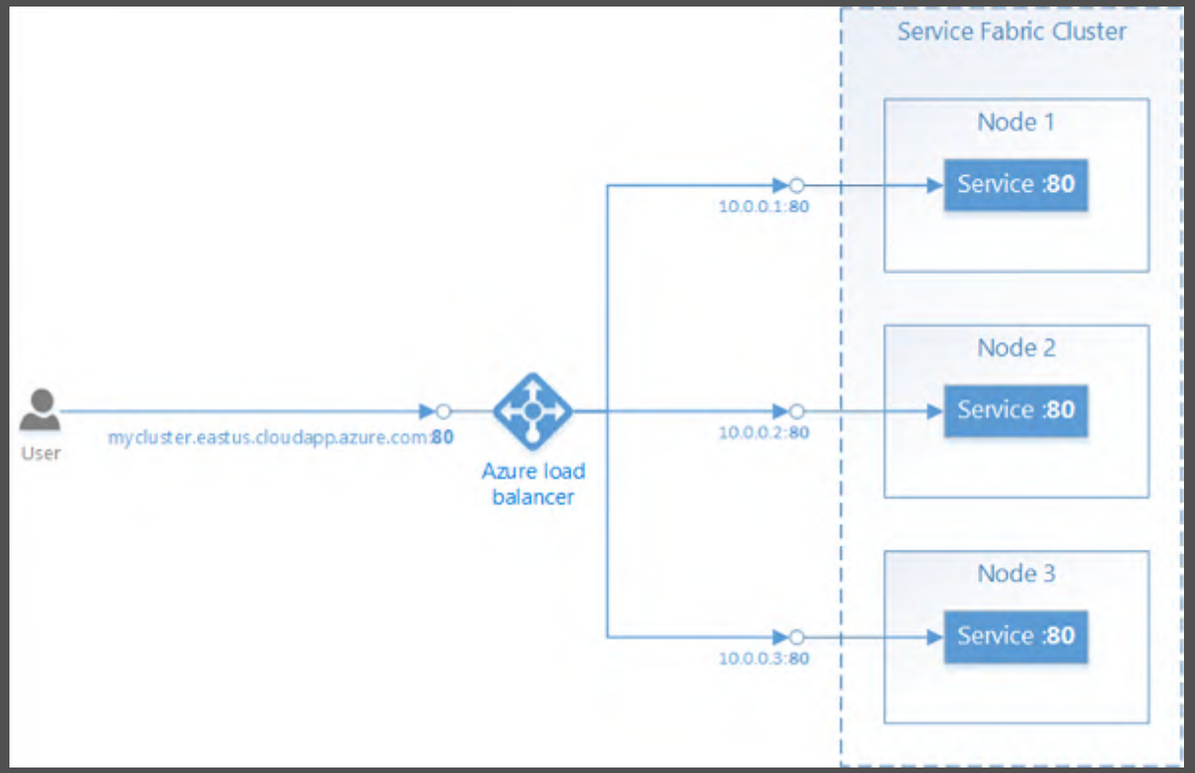


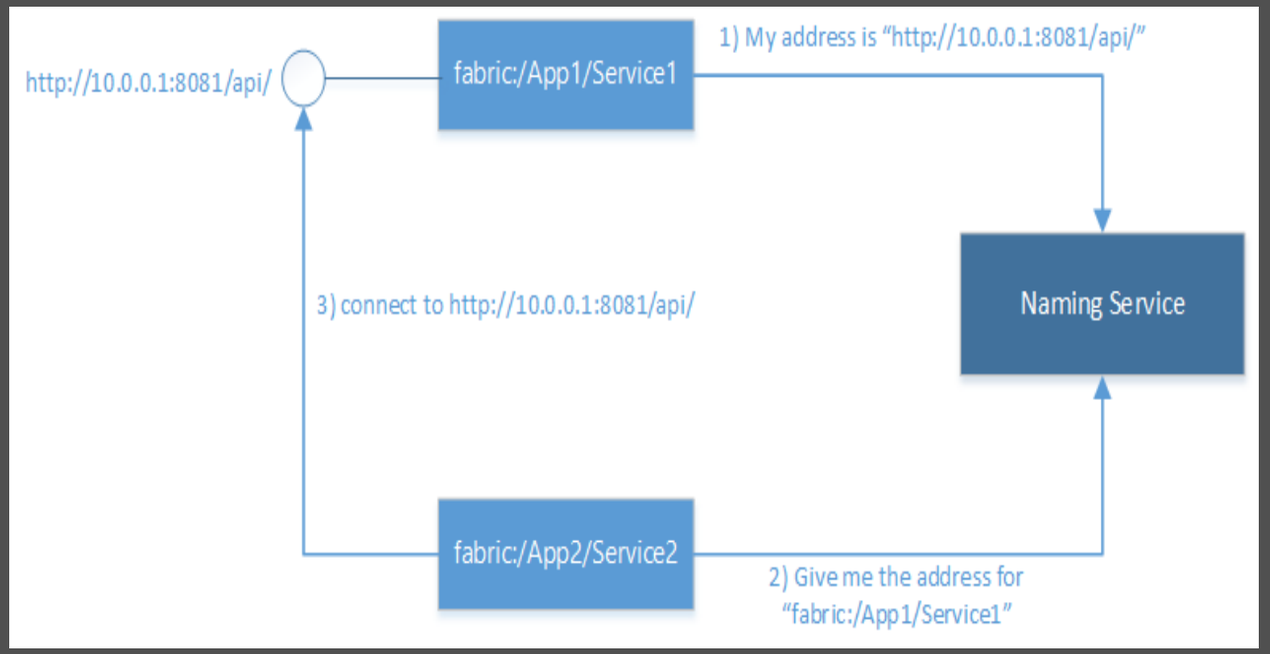
Service Fabric –
底层运维支撑体系

分区以高并发，副本以高可用



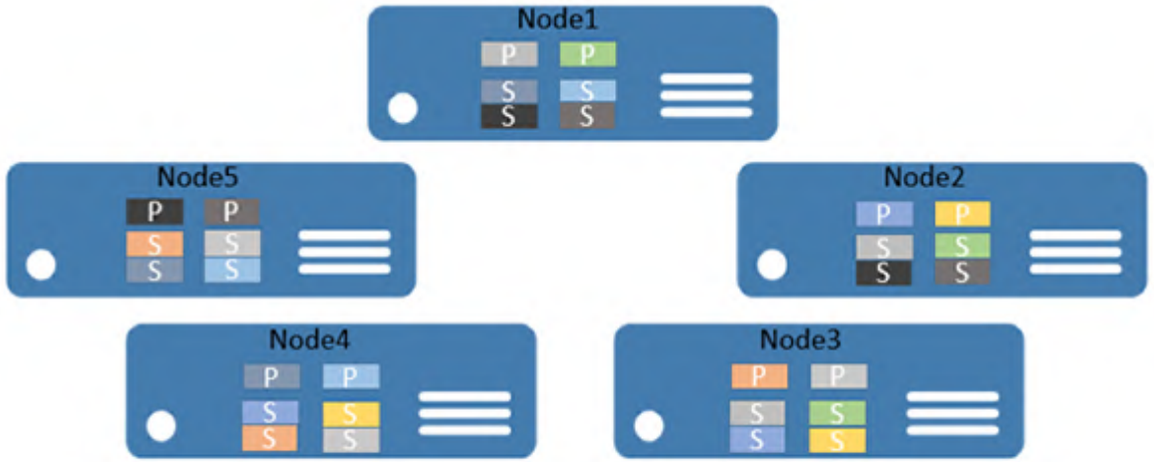
Naming Service 发现和负载均衡 “IP + Port 服务实例”



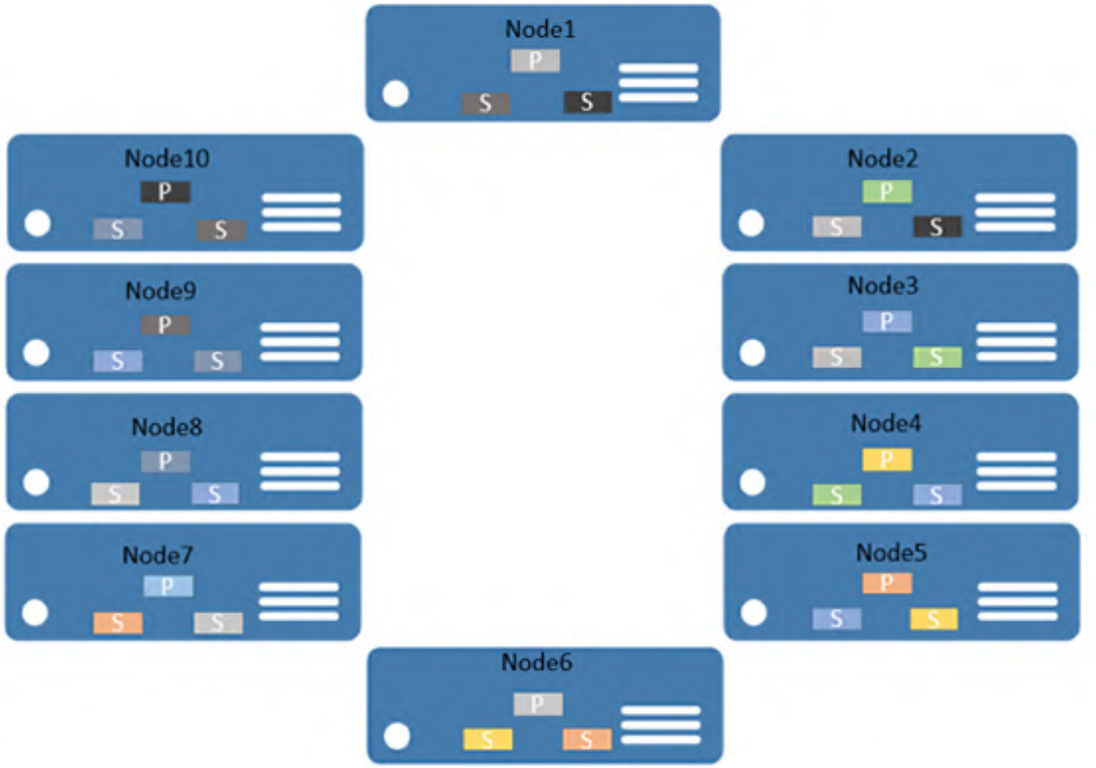


微服务实例 在节点集群下的自动部署

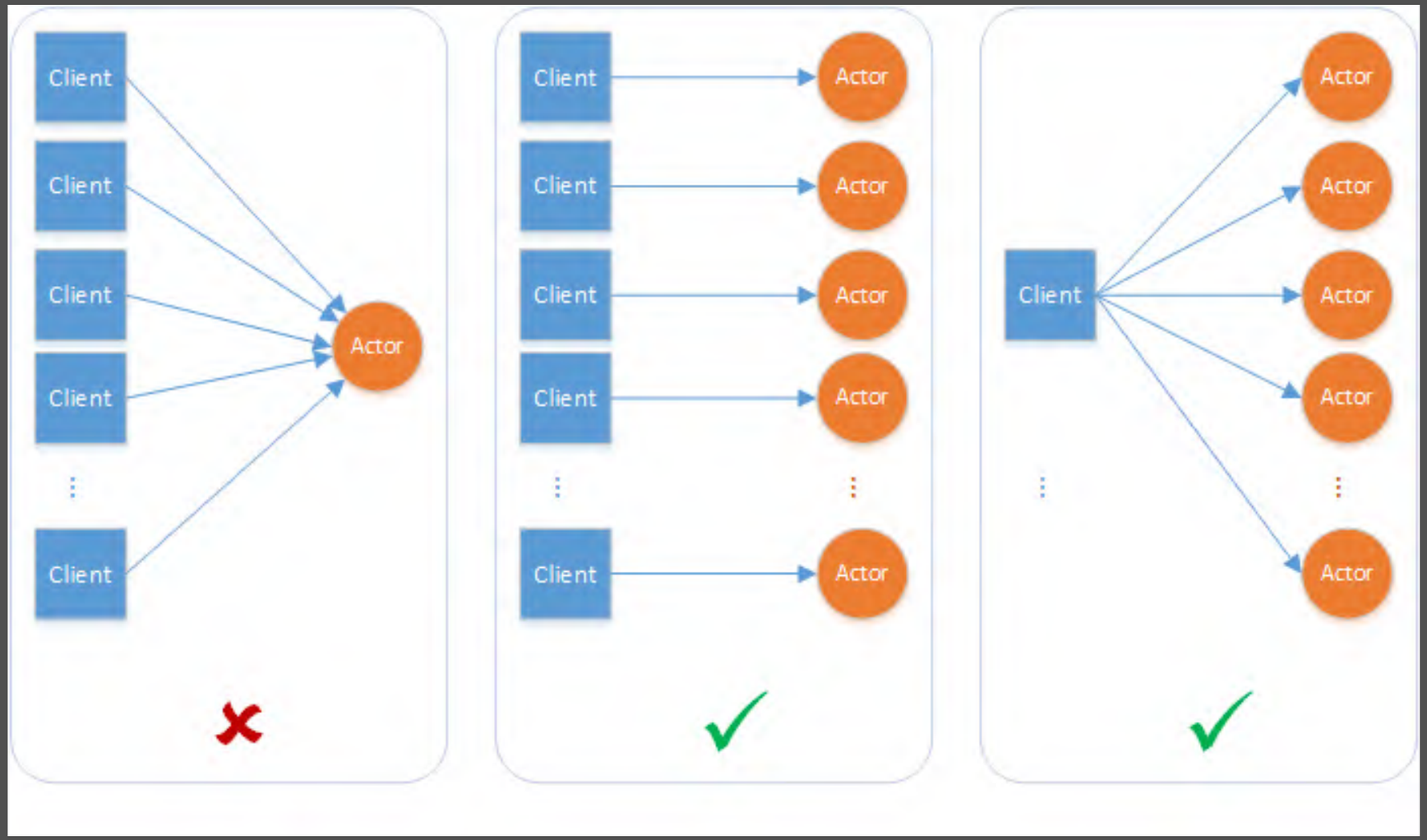
5-node cluster with 10 partitions



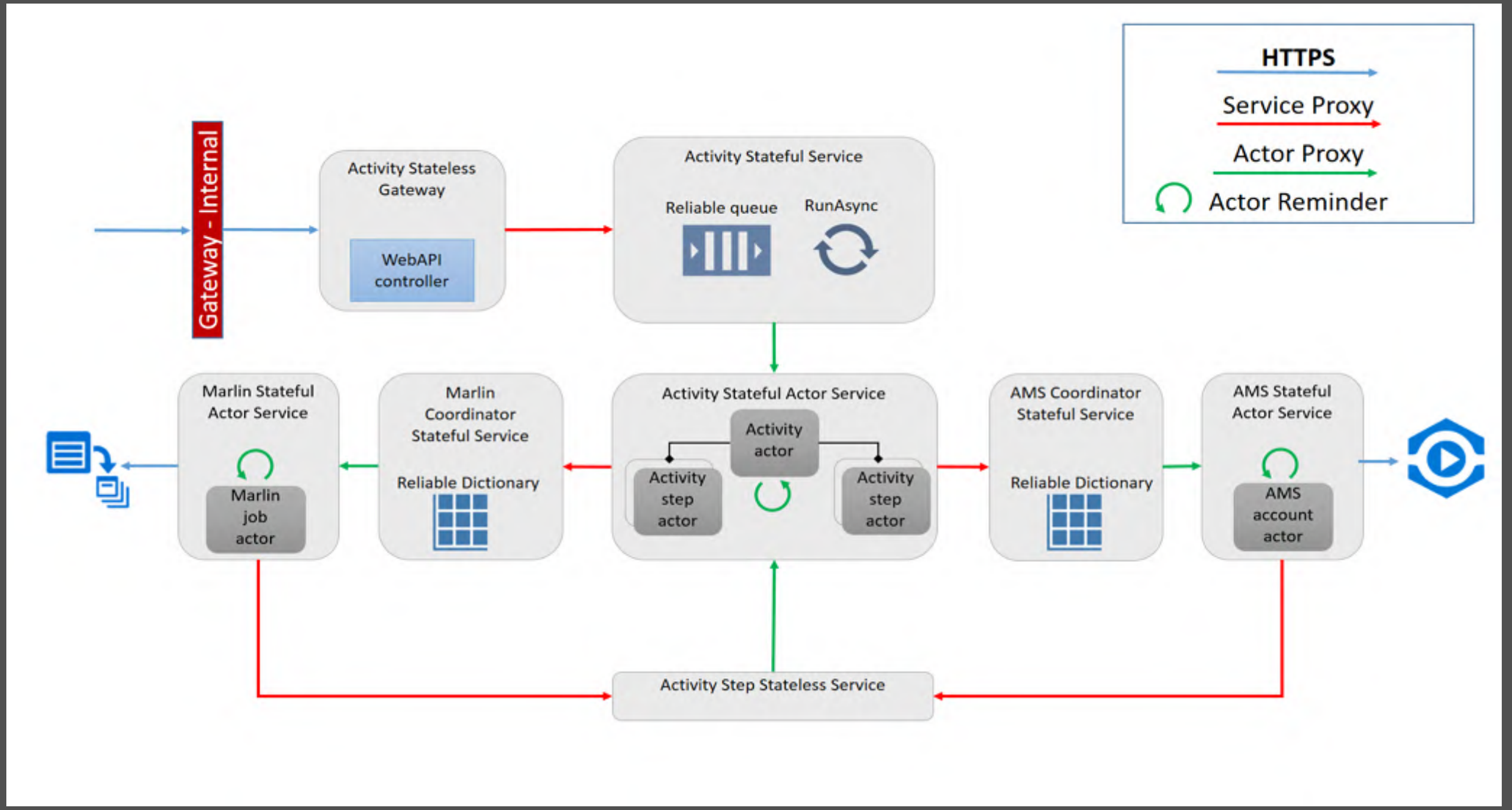
10-node cluster with 10 partitions



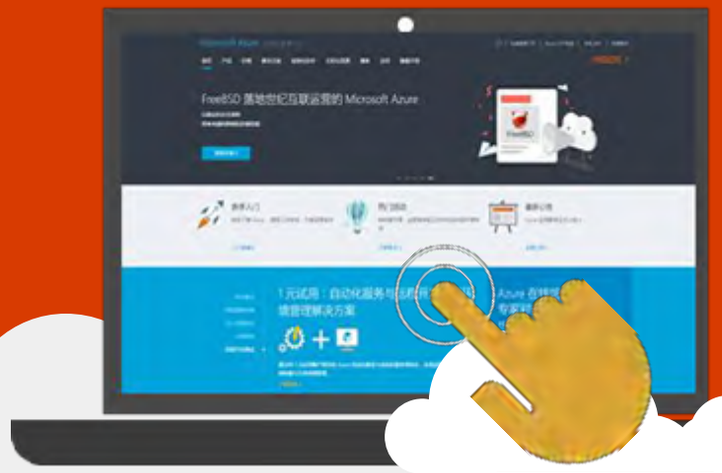
Actor - 单线程模式代理， 类似内存管理的持久化运行



一个微服务改造的应用例子



更多信息与资源



- Azure 中国官网 https://www.azure.cn/ 提供最新产品与解决方案信息， 技术文档，以及SDKs下载
- Azure 应用程序开发说明 https://www.azure.cn/dev-notes/ 概述了海外与中国区服务开发人员需要注意的区别
- 申请一元试用，即刻体验 Azure 服务：https://www.azure.cn/pricing/1rmb-trial-full/
- Azure 镜像市场： https://market.azure.cn/

Microsoft 云科技公众号

Azure 云助手手机 App

Thank you!