

# Docker与MongoDB云服务

罗成对

luochengdui@ucloud.cn

QQ: 376925324

# 自我介绍

- 负责UCloud UDB产品研发、运营
- 浙江大学毕业后就职于华为中央研究院、UCloud
- 从事分布式系统、数据库、NoSQL以及云计算平台的研发
- 推出国内第一家MongoDB云服务

# 目录

- 云数据库服务
- 为什么使用Docker
- 分布式集群Docker化
- 实践效果

# 用户眼中的云数据库

- 高可用和高性能
- 按需
  - 按需付费
  - 按需扩展
- 可控和可运维
  - 希望拥有权限
- 用户体验
  - 一键部署、秒级交付

# 我们眼中的云数据库

- 多租户
  - 用户行为不可预知
- 资源整合
  - 分配与隔离
- 低成本
  - 充分利用资源
- 可控和可运维
- 可移植

# 为什么使用Docker

- 轻量级虚拟化
  - LXC
  - Control Groups
  - Linux Namespace
- 微服务设计
  - 服务原子化
  - 解耦与重组
- VM or Docker?
  - 选择Docker，比虚拟机占用了更少的资源，比物理机隔离性更好

# 优势互补

Feature	Docker	MongoDB
High Performance	√	√
Scale-up	√	-
Scale-out	-	√
High Availability	-	√
Cost	√	-
Deployment/Maintenance	√	-
Security	√	-

# 分布式集群Docker化

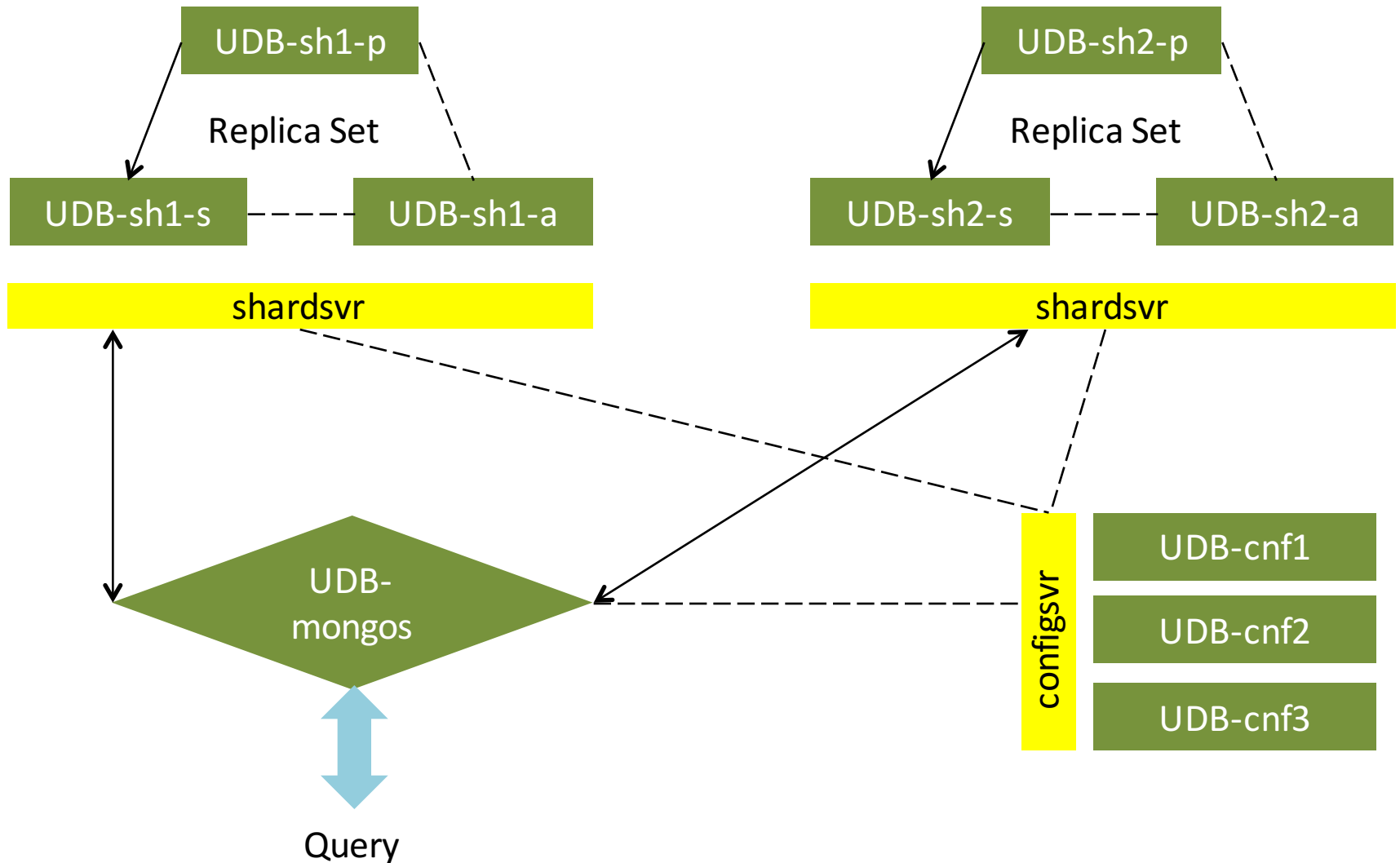
- 集群介绍
- 集群架构和部署
- 生产运维
- 诊断与监控



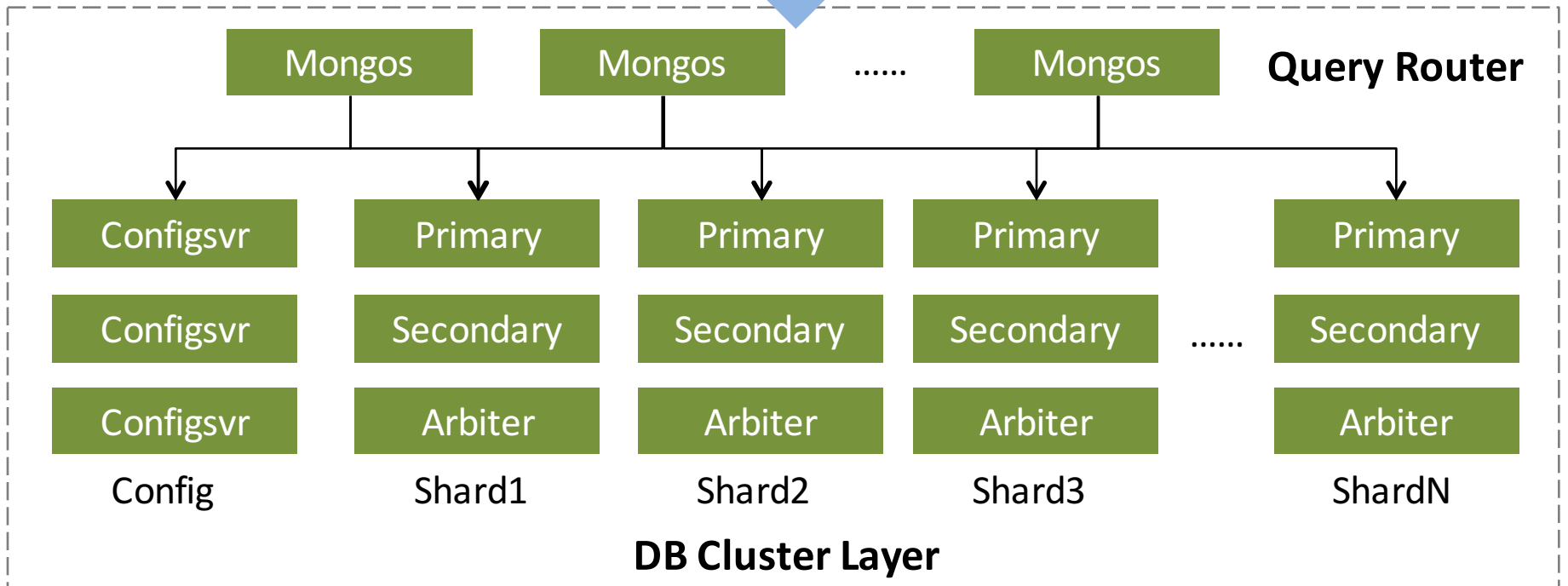
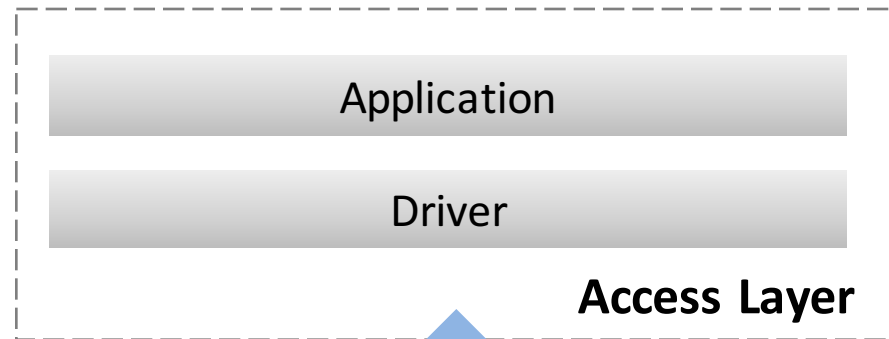
# 应用场景

- 数据源及规模
  - 类型：业务类、日志类、监控类、元数据...
  - 特点：增长快、schema多样
  - 规模：轻松达到单表TB级
- 面临的问题及对策
  - 单点压力（热点库表）--scale-up和scale-out, high performance
  - 单点故障 --HA
  - 成本 -- 按需申请资源
  - 运维 -- 易管理、易部署
- 高可用、高扩展性的云数据库MongoDB集群

# 弹性MongoDB集群



# 弹性MongoDB集群



# 控制台

北京BGP C

消息

帮助

luochengdui@ucloud.cn

+ 创建MongoDB实例

全部



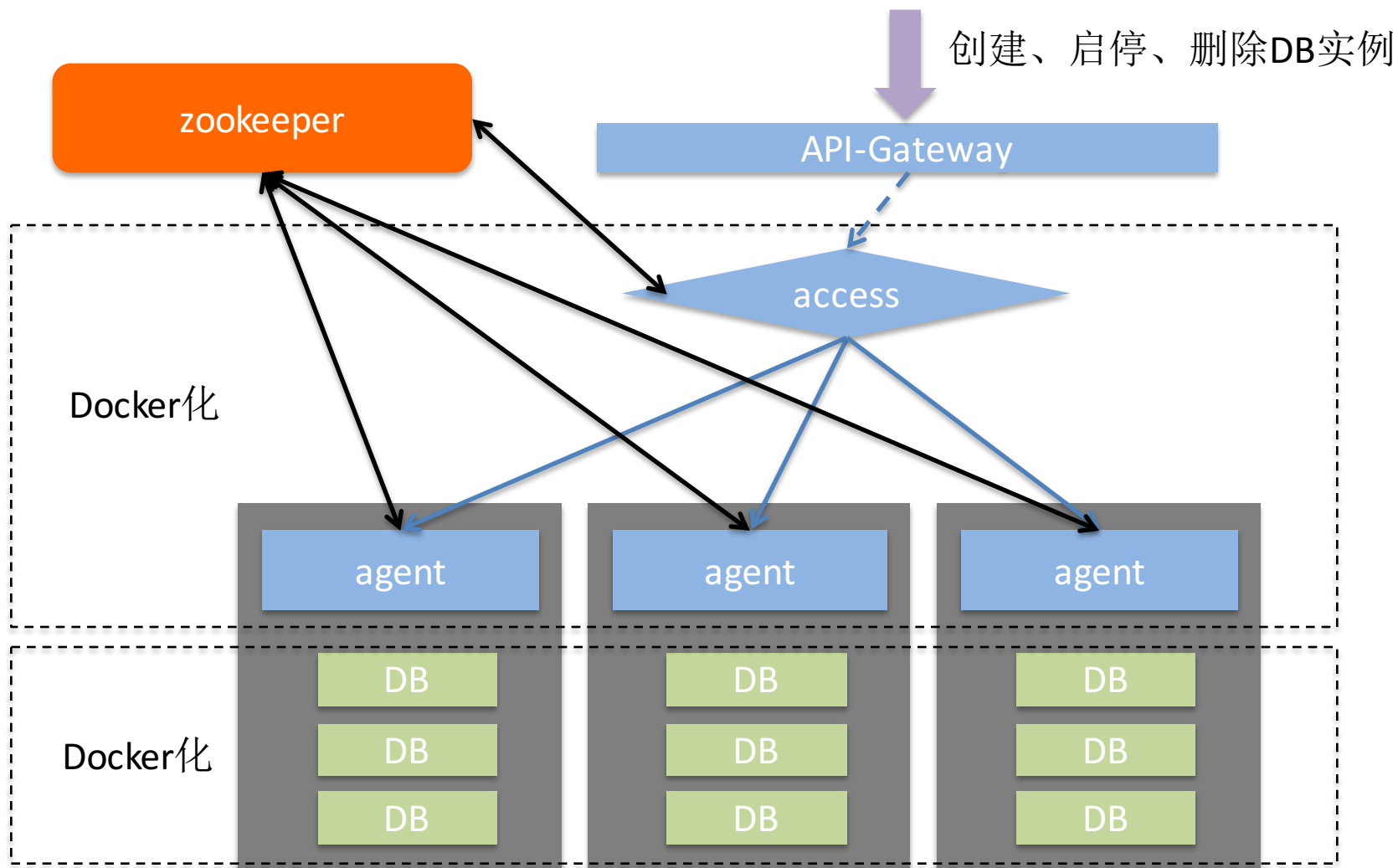
实例名称	属性	IP地址	类型	下次扣款日	状态	操作
production-config	configsvr	10.10.128.111	mongodb-2.6	2015-08-21 09:52:14	运行	
production-config	configsvr	10.10.142.161	mongodb-2.6	2015-08-21 09:52:14	运行	
production-config	configsvr	10.10.138.51	mongodb-2.6	2015-08-21 09:52:14	运行	
production-mongos1	mongos	10.10.114.53	mongodb-2.6	2015-08-21 23:21:25	运行	
production-mongos2	mongos	10.10.121.139	mongodb-2.6	2015-08-21 23:21:55	运行	
production-shard1-primary	primary	10.10.137.235	mongodb-2.6	2015-08-21 09:50:48	运行	创建节点
production-shard1-secondary	secondary	10.10.115.160	mongodb-2.6	2015-08-21 09:53:02	运行	
production-shard1-arbiter	arbiter	10.10.120.75	mongodb-2.6	2015-08-21 09:54:02	运行	
production-shard2-primary	primary	10.10.114.215	mongodb-2.6	2015-08-21 09:51:40	运行	创建节点
production-shard2-arbiter	arbiter	10.10.116.161	mongodb-2.6	2015-08-21 09:55:47	运行	
production-shard2-secondary	secondary	10.10.128.103	mongodb-2.6	2015-08-21 09:53:15	运行	
production-shard3-primary	primary	10.10.120.82	mongodb-2.6	2015-08-21 09:51:40	运行	创建节点
production-shard3-secondary	secondary	10.10.118.205	mongodb-2.6	2015-08-21 09:53:29	运行	
production-shard3-arbiter	arbiter	10.10.128.235	mongodb-2.6	2015-08-21 09:54:33	运行	

configsvr

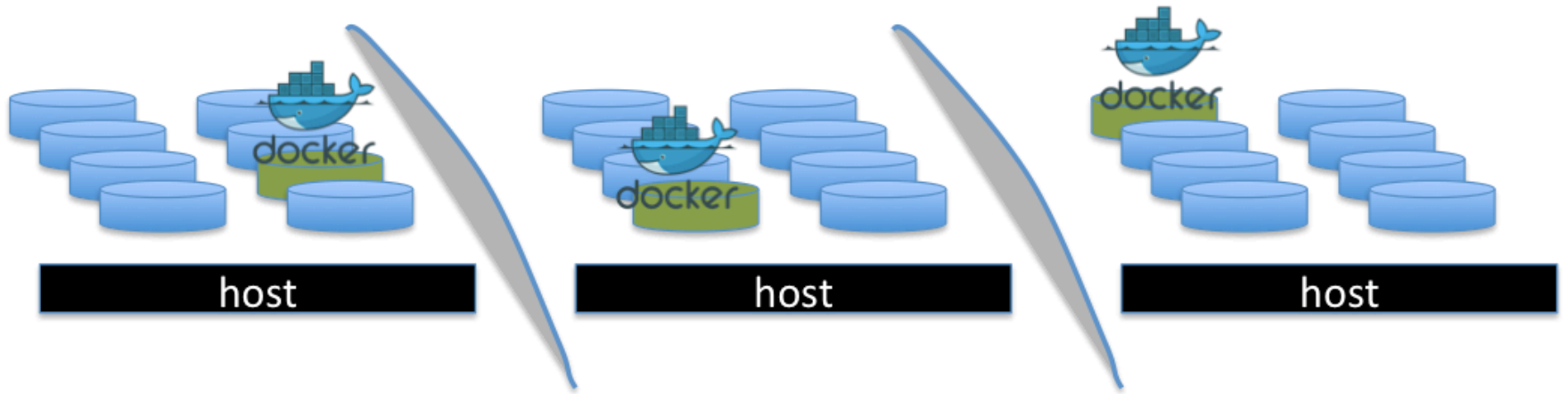
mongos

shardsvr

# 架构

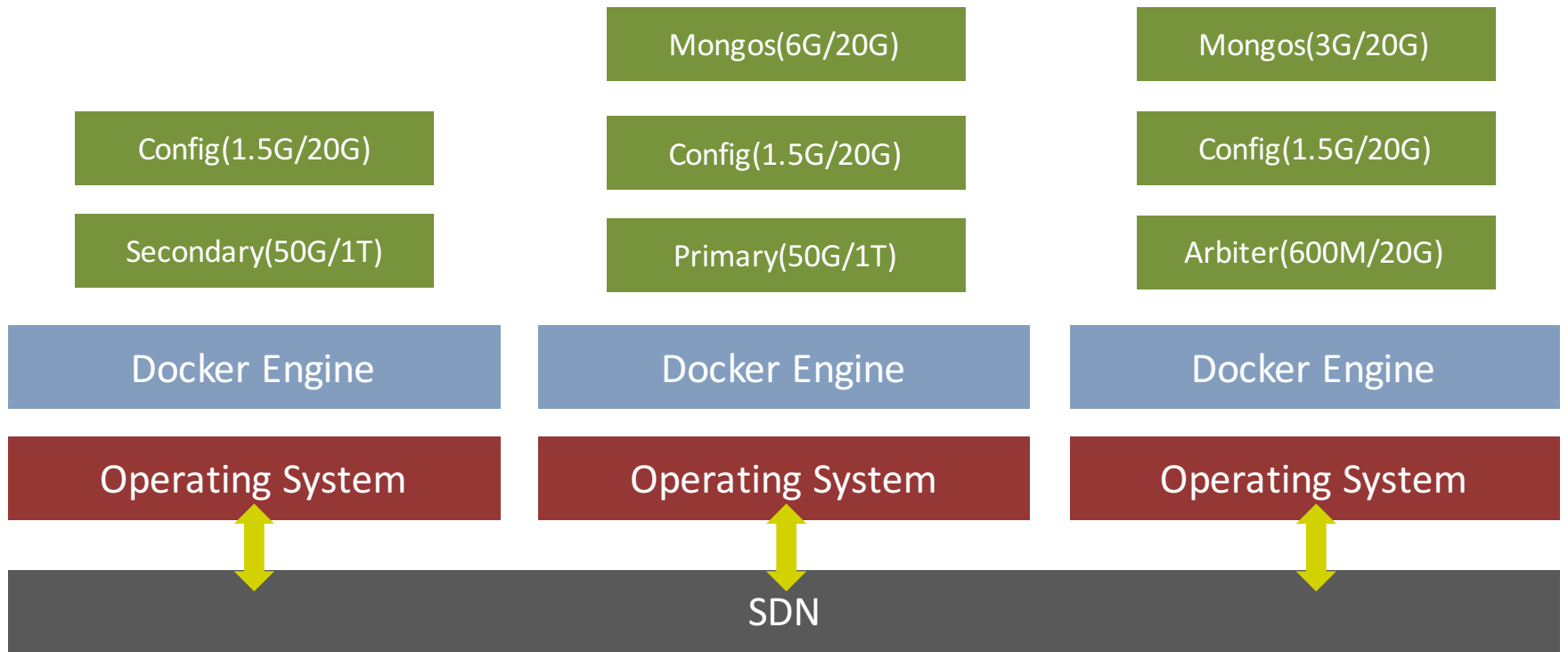


# 部署



# 部署

## On-Demand/Fast/Auto-Scheduling



# Docker化的挑战

- 数据库有状态
  - 数据卷挂载
  - 多目录映射
  - 离线迁移或者在线迁移
- SDN环境网络通信
  - 自定义网桥
  - 虚拟网络设备veth
- 管理模块（access、agent等）多版本



# 部署

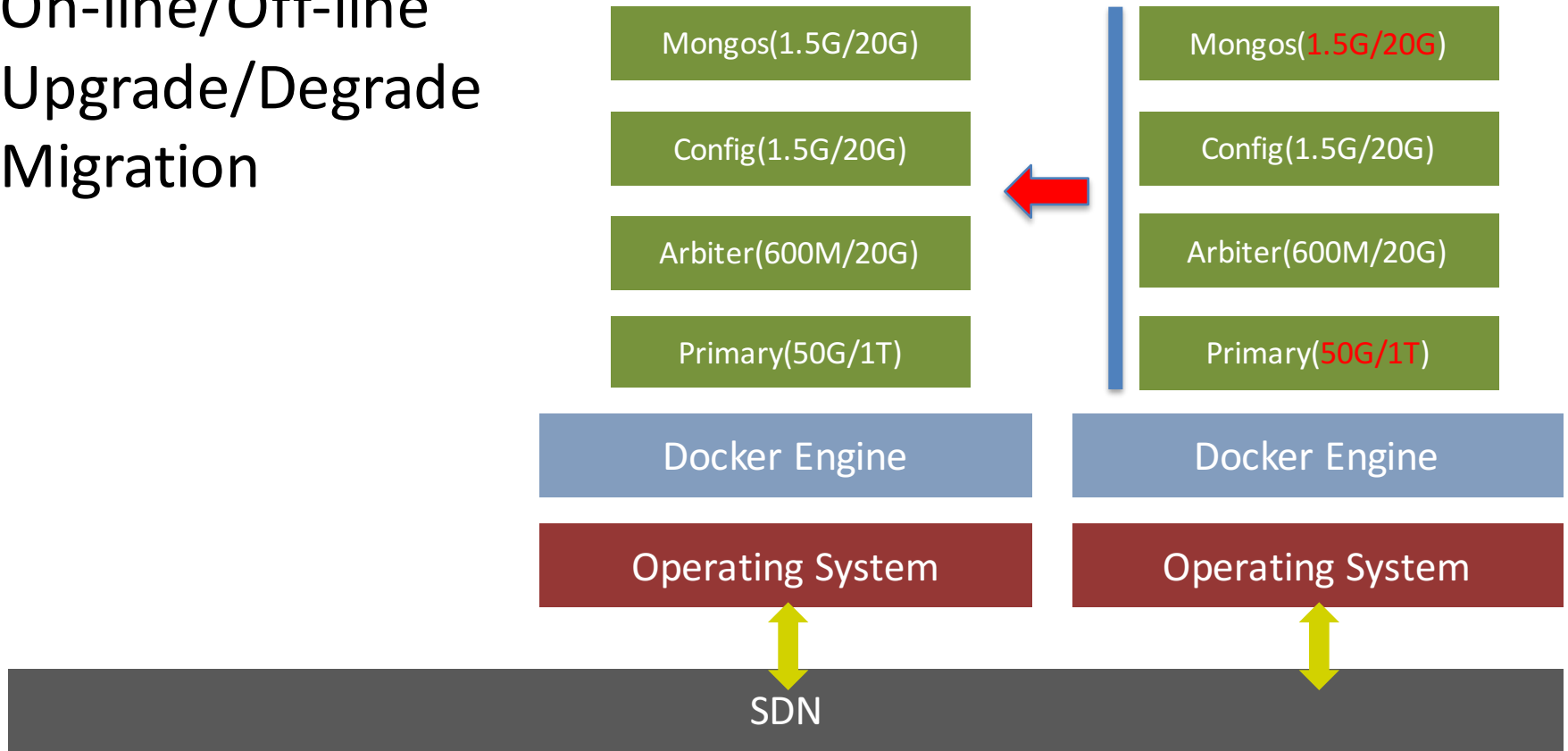
- Image
  - 基础Image来源干净可靠（推荐官网）
  - 利用Dockerfile安装必要工具链
  - 合理的版本迭代
- 操作系统
  - 内核版本和Docker适配
  - Cgroup挂载
  - 最大文件数，限制DB的连接数等等
- 数据卷持久化数据
  - 设置合理的读写权限

# 部署

- 网络设置
  - 自定义网桥，限制docker0
  - 内网隔离
  - 禁止暴露端口
  - 配置DNS等
- 安全加强
  - 启用SELinux/GRSEC
  - 启用能力机制，注意某些超级权限
- Docker Daemon防护
  - 禁止宿主机根目录映射
  - 禁止滥用root权限

# 运维

On-line/Off-line  
Upgrade/Degrade  
Migration



# 异常现象剖析

- Service异常
  - 现象： 停服、自动重启， 概率极低
  - 后果： Container存活状态存在不确定性
    - （1） killed（时间上不确定） → 无法umount挂载点 → 无法启动Container
    - （2） hang住
  - 措施： 实时监控Docker Service
- 数据卷读写异常
  - 现象： rw变ro， 小概率事件
  - 后果： 映射目录无法写入
  - 措施： 重建目录挂载

# 异常现象剖析

- Docker版本升级异常
  - 现象：向下不兼容，概率不确定，版本相关
  - 后果：已存images消失，后果很严重
  - 措施：预防，保持存量机器不升级
- 最大文件数
  - 现象：默认1024，限制DB连接数
  - 后果：DB无法连接
  - 措施：
    - (1) 为Docker Daemon设置合理的ulimit
    - (2) 更改父进程ulimit

# 异常现象剖析

- 低版本Docker触发bug
  - 现象：容器无法关闭、正常启停，更严重时宿主机agent服务异常
  - 后果：DB不可用、宿主机各类服务不可用
  - 措施：内核热补丁、Docker热升级

# 监控与诊断

- 监控级别
  - 平台级监控和拨测
  - 实时监控宿主机存活、网络连通性
  - 实时监控Docker Service
  - 实时监控Container存活
  - 实时监控DB连通性
- Log
  - 宿主机日志/var/log/messages
  - Docker日志/var/log/docker
  - DB系统日志
  - 业务逻辑日志

# 实践效果

- 效果
  - 高性能，低损耗
  - 按需申请，秒级安装
  - 一键式集群部署
  - 容器稳定可靠
- 现状
  - 已有几千个Docker容器在运行
  - Docker化的实例包括agent实例和DB实例





Thank You