DevOps in the Cloud

茹云峰
@fengyuncrawl
2016.11.18

about me

c，php，python，go，erlang

热衷于搜索，社交，数据挖掘，系统架构

DevOps

服务发现（api gateway）

docker

集群资源管理

调度系统

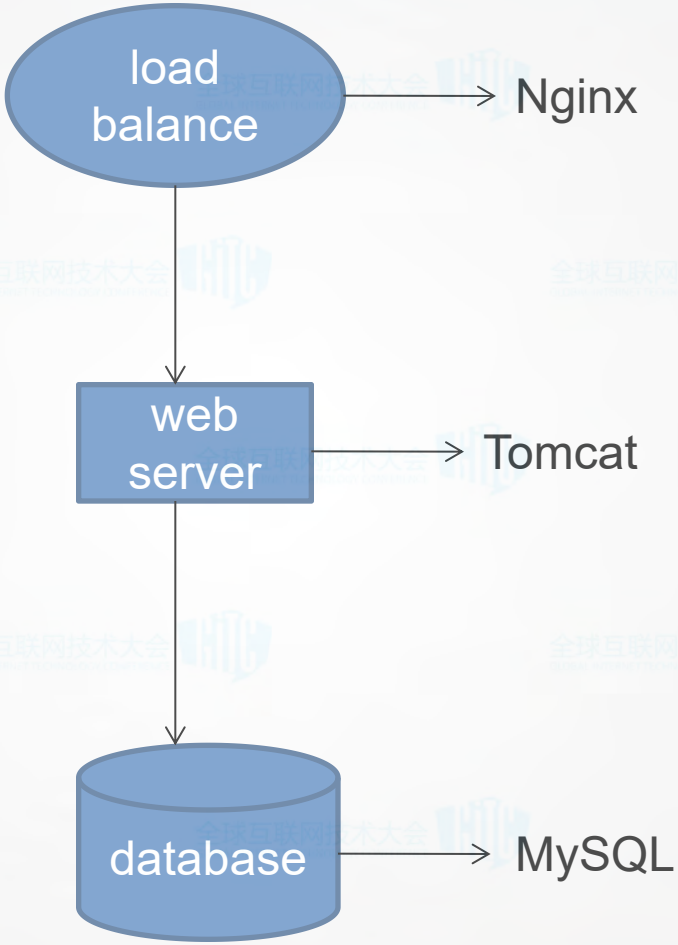
分布式tracing

日志搜索

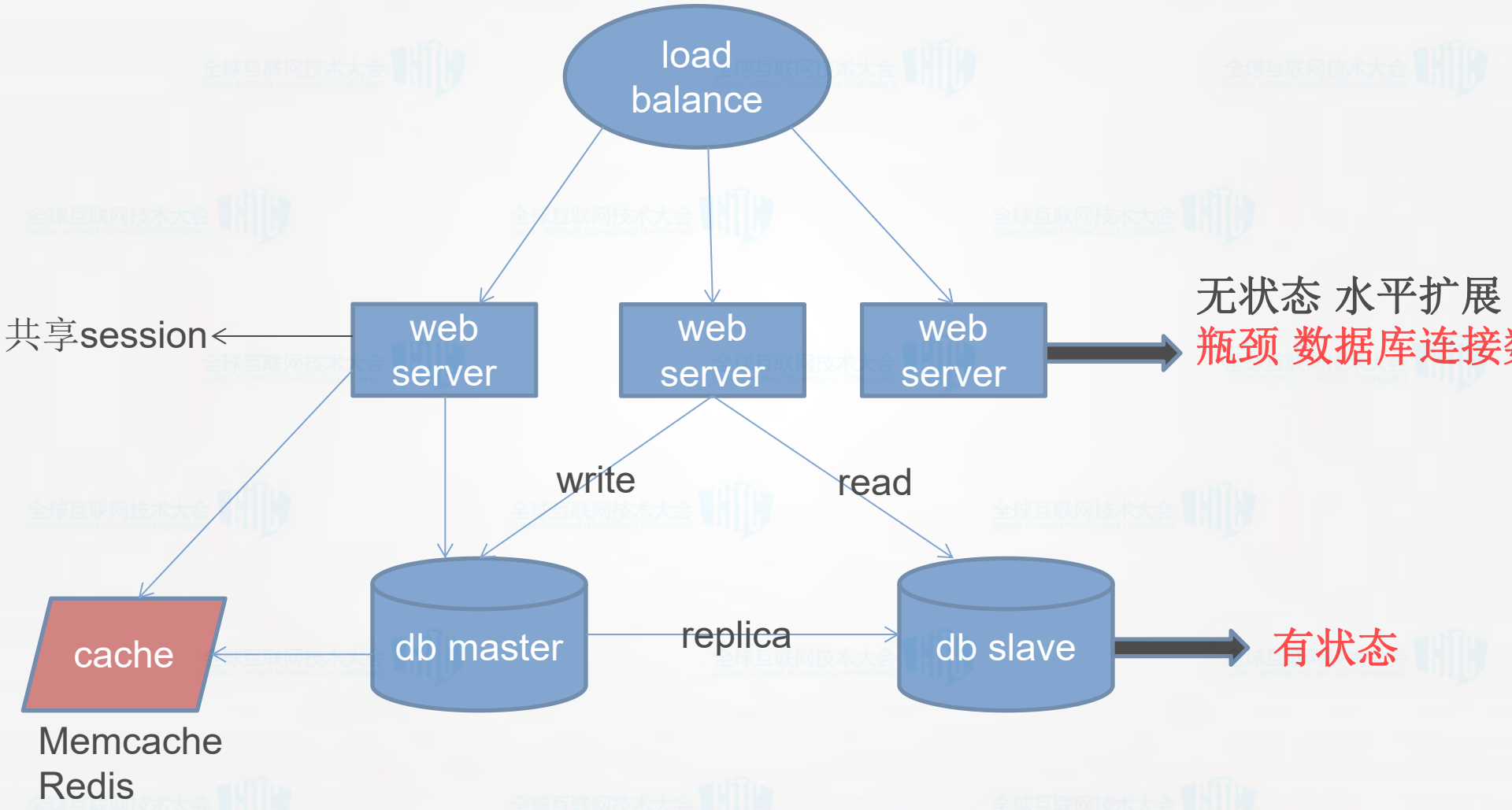
metric性能监控

autoscaling

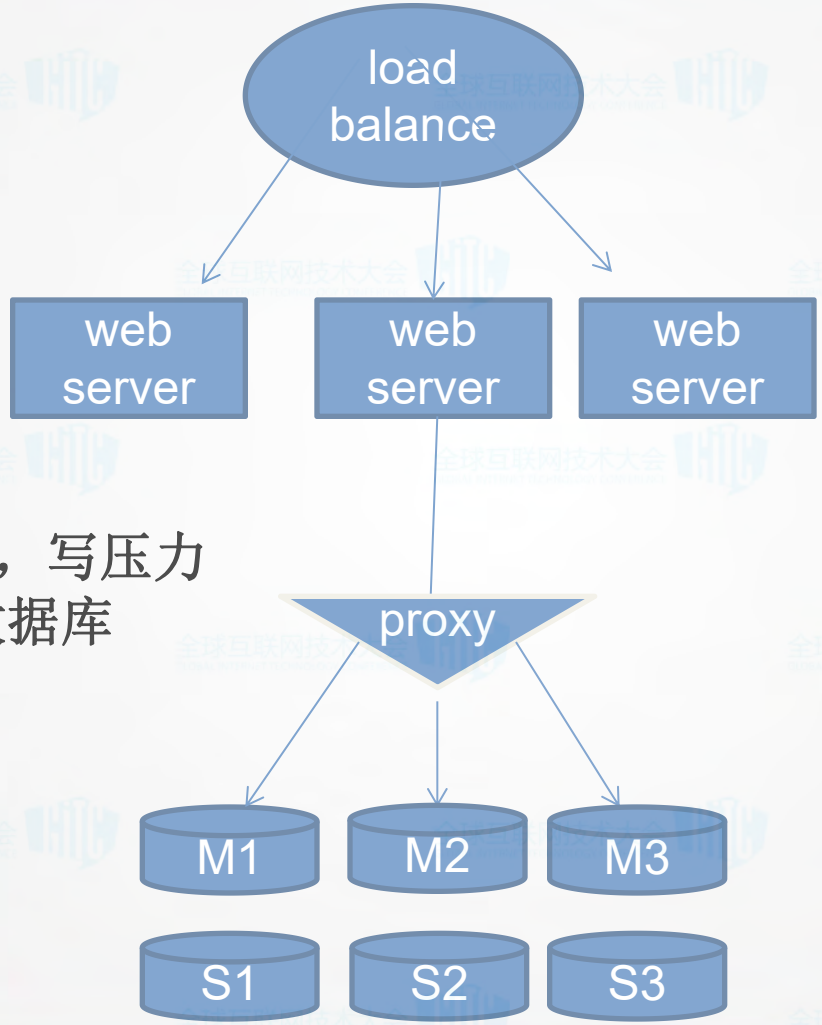
单机时代



多机分布


load
balance
web
server
web
server
web
server
共享session
无状态 水平扩展
瓶颈 数据库连接数
write
read
cache
db master
replica
db slave
有状态
Memcache
Redis

日志监控

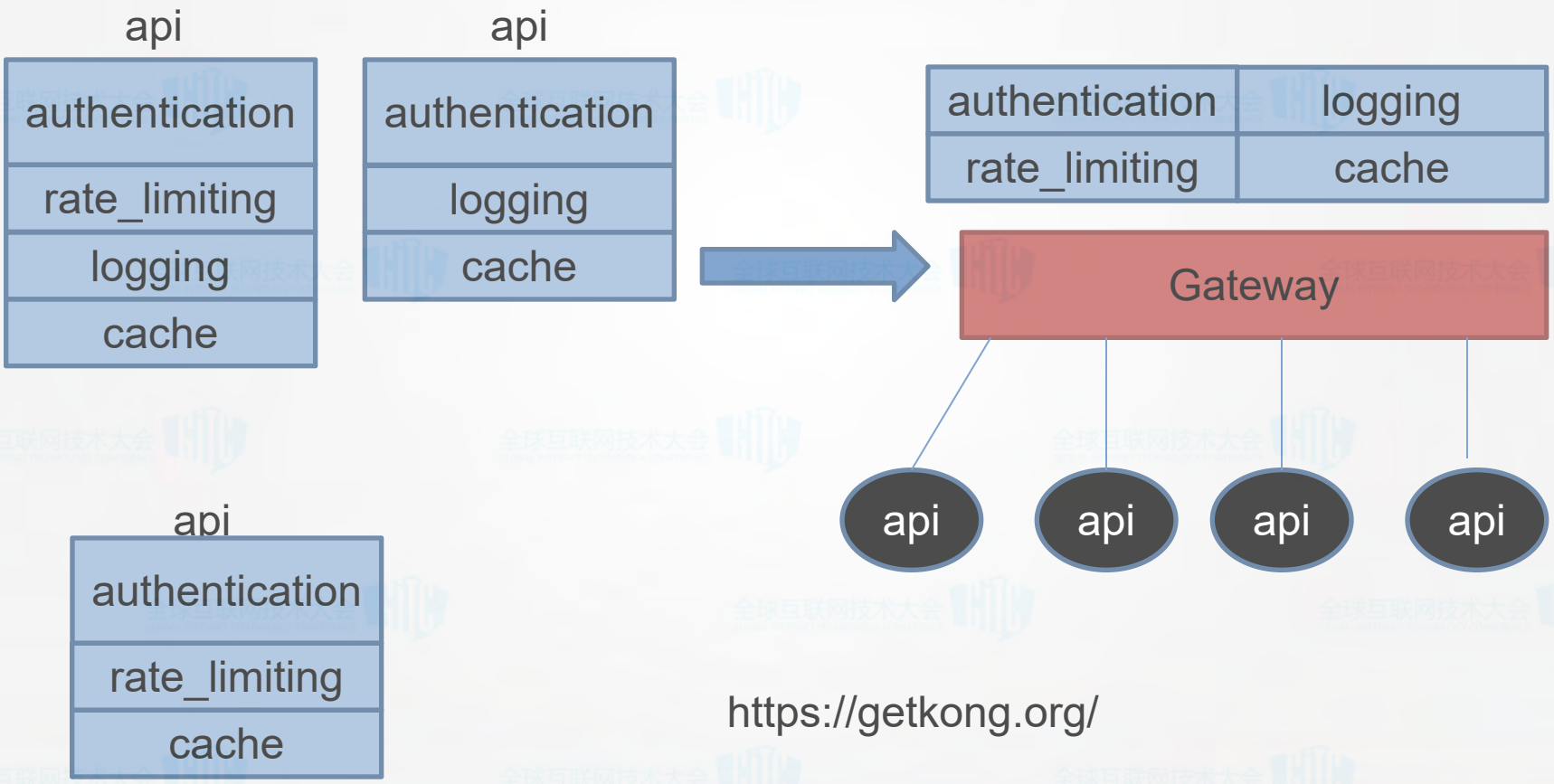


数据库有状态，写压力大 只能切分数据库

Elasticsearch+Logstash+Kibana

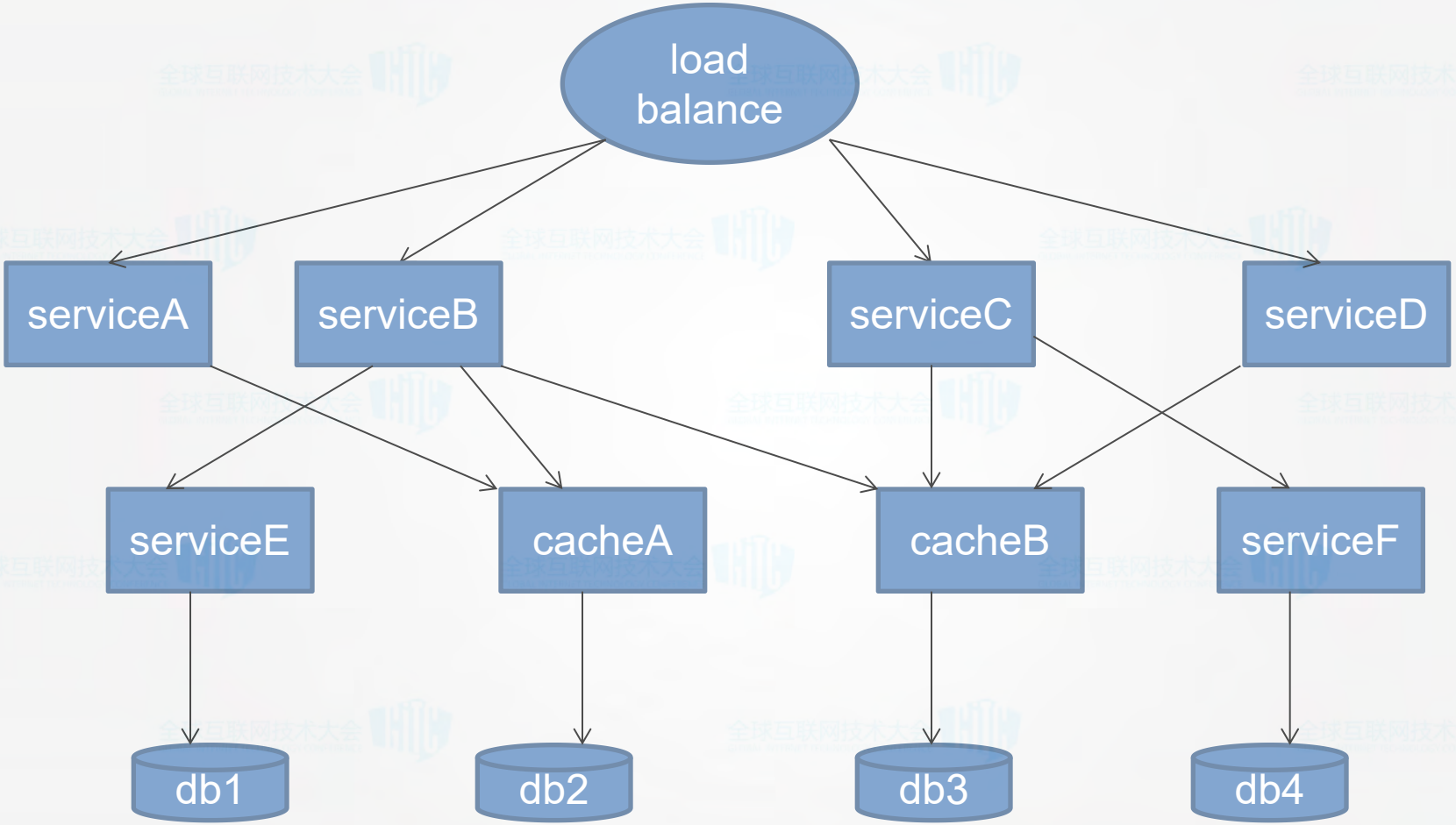
logstash性能差
替换成Rsyslog或者Nxlog

API Gateway



https://getkong.org/

RPC


load
balance
serviceA
serviceB
serviceC
serviceD
serviceE
cacheA
cacheB
serviceF
db1
db2
db3
db4

问题

Api太多管理复杂

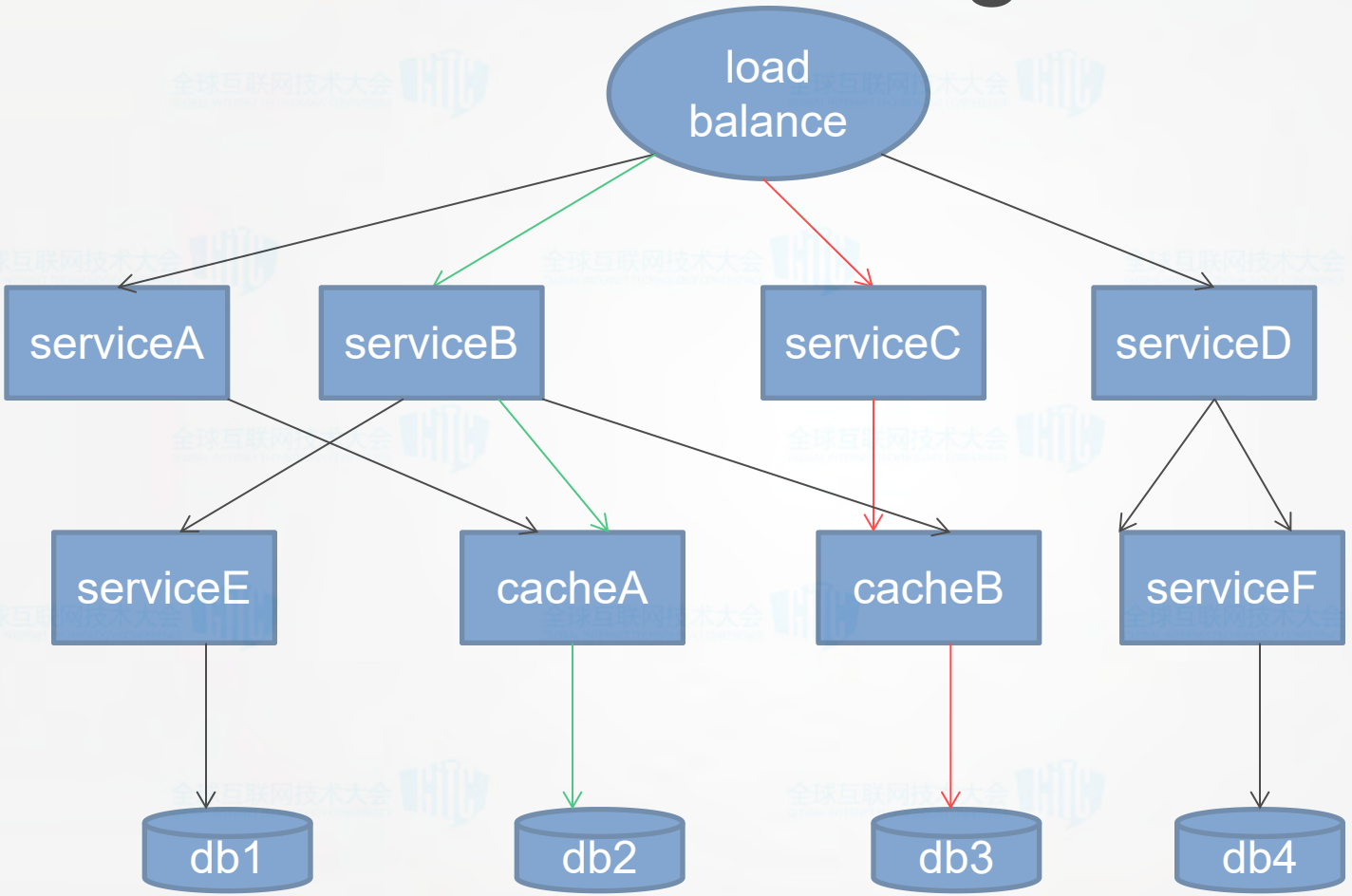
service之间调用关系复杂（A->B->C）

压力测试复杂

部署复杂

无法追踪系统问题

分布式tracing/APM



Pinpoint
java

phpio/phptrace
php

open
tracing

挑战

能否替代单机版MySQL？

能否放弃数据库中间件和分库分表，并且保证跨行跨表事务？

像单机一样操作分布式关系型数据库

oltp

Tidb（测试中）

Trafodion（运维复杂）

CockroachDB(可用性不行)

trafodion

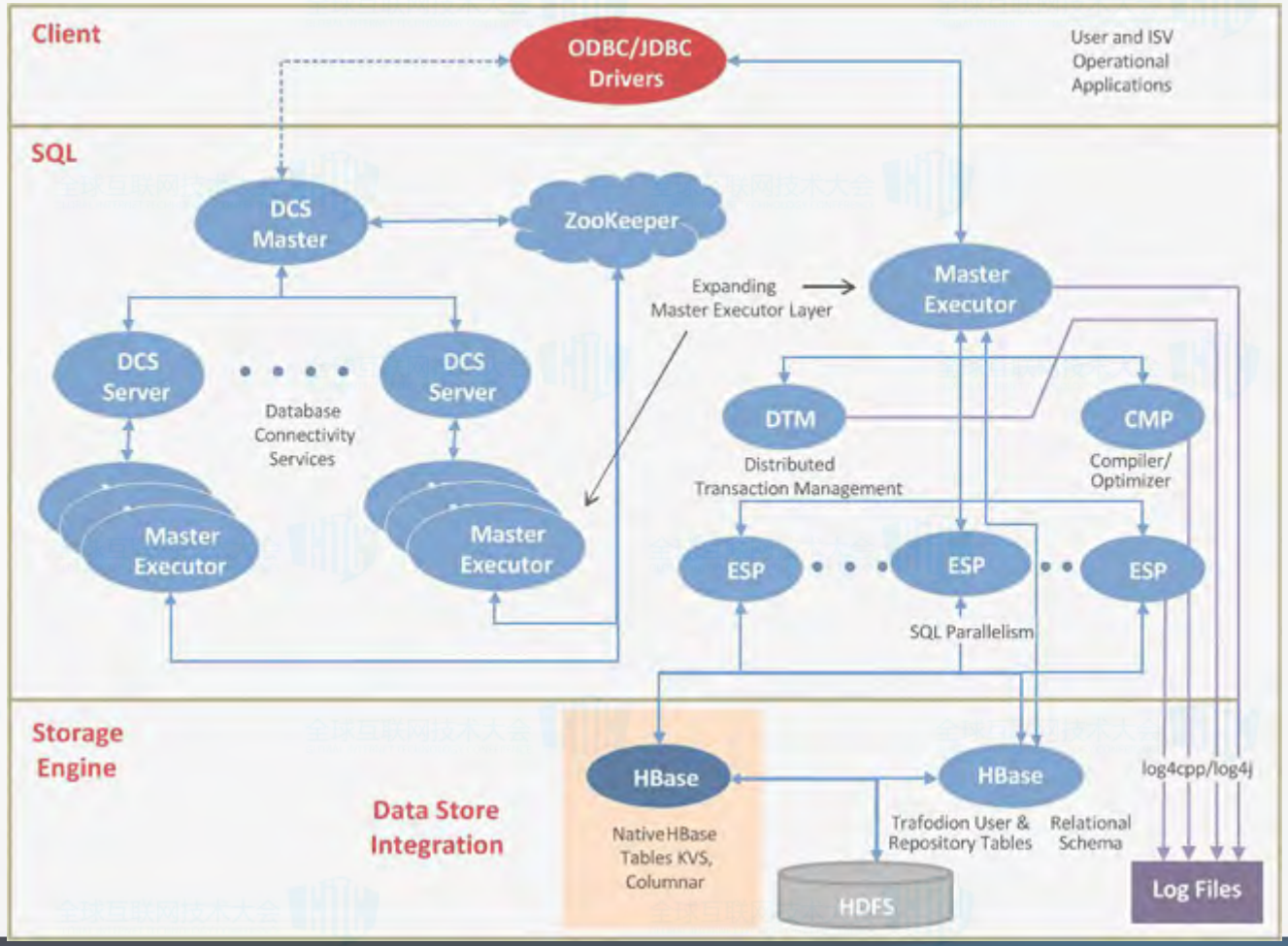
基于hbase/hadoop分布式关系型数据库

实时高并发

分布式跨行跨表ACID

建议安装2.0.0最新版以及hbase1.0

缺点：运维复杂



olap

ClickHouse　https://clickhouse.yandex

列式存储，实时分析，distributed joins，restful api，跨数据中心复制

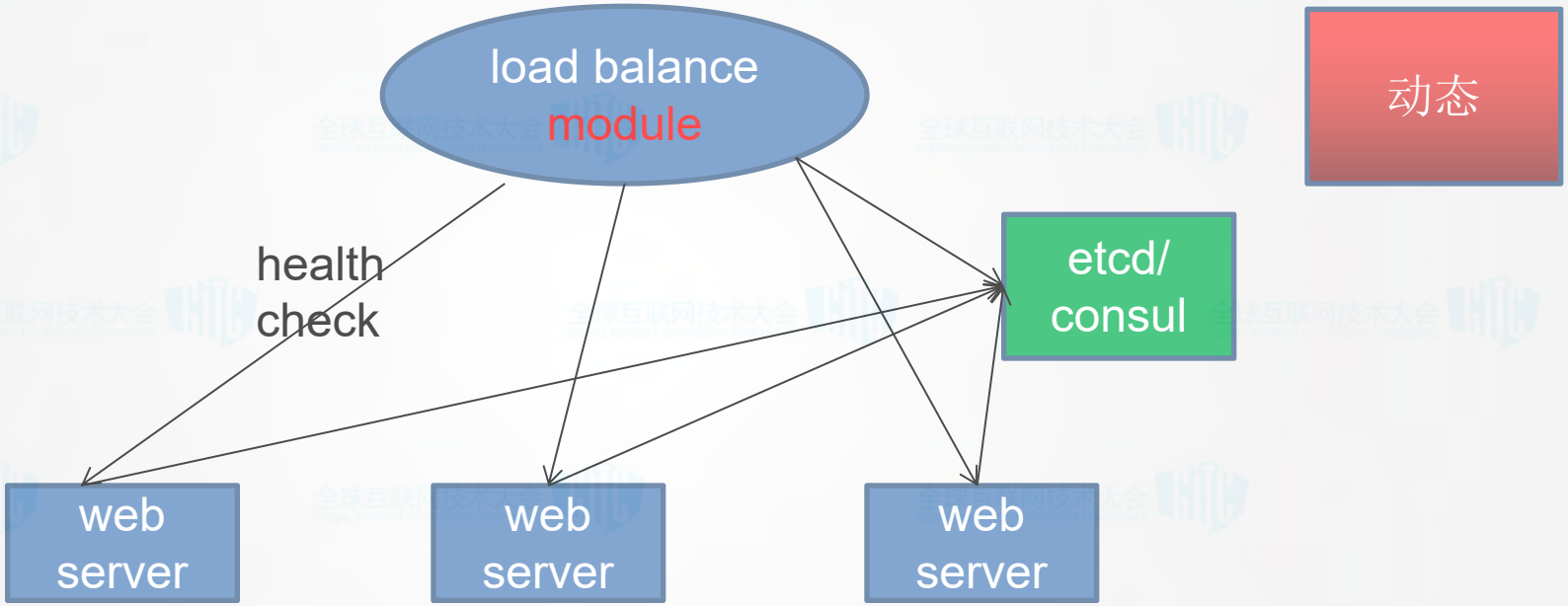
Druid(聚合)

Apache Kylin（用空间换时间）

Kudu（兼有hbase and HDFS）

问题

上述架构的缺点：Load balance与后端web server强耦合，如果流量突发，需要经常修改config ip router



https://github.com/weibocom/nginx-upsync-module

docker优缺点

优点：

进程级（轻量级）虚拟化

快速部署

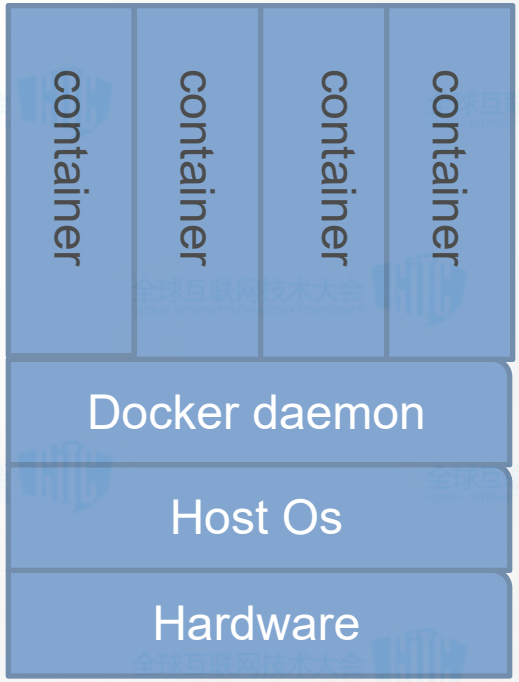
缺点：

多租户安全性

网络

文件系统

有状态服务（数据库）不完善

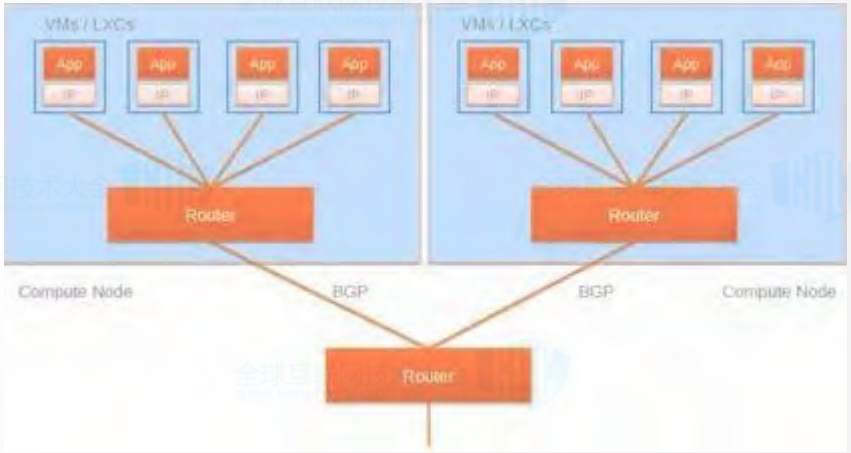


docker网络

host

bridge

container



Calico

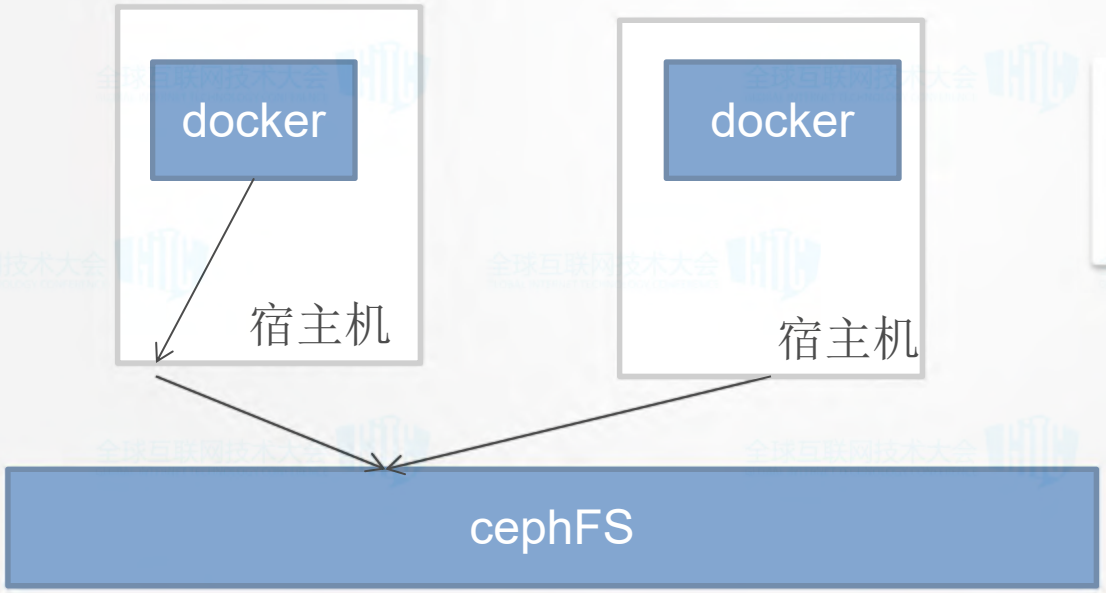
OpenVSwitch（配置复杂）

Weave，Flannel（性能差）

docker文件系统

挂载宿主机本地文件系统做持久化

挂载分布式块存储做共享持久化



有状态服务（数据库）
状态+存储是个挑战
如何高可用，如何分布式存储？

CI/CD

访问控制

审计日志

web admin ui

https://github.com/vmware/harbor

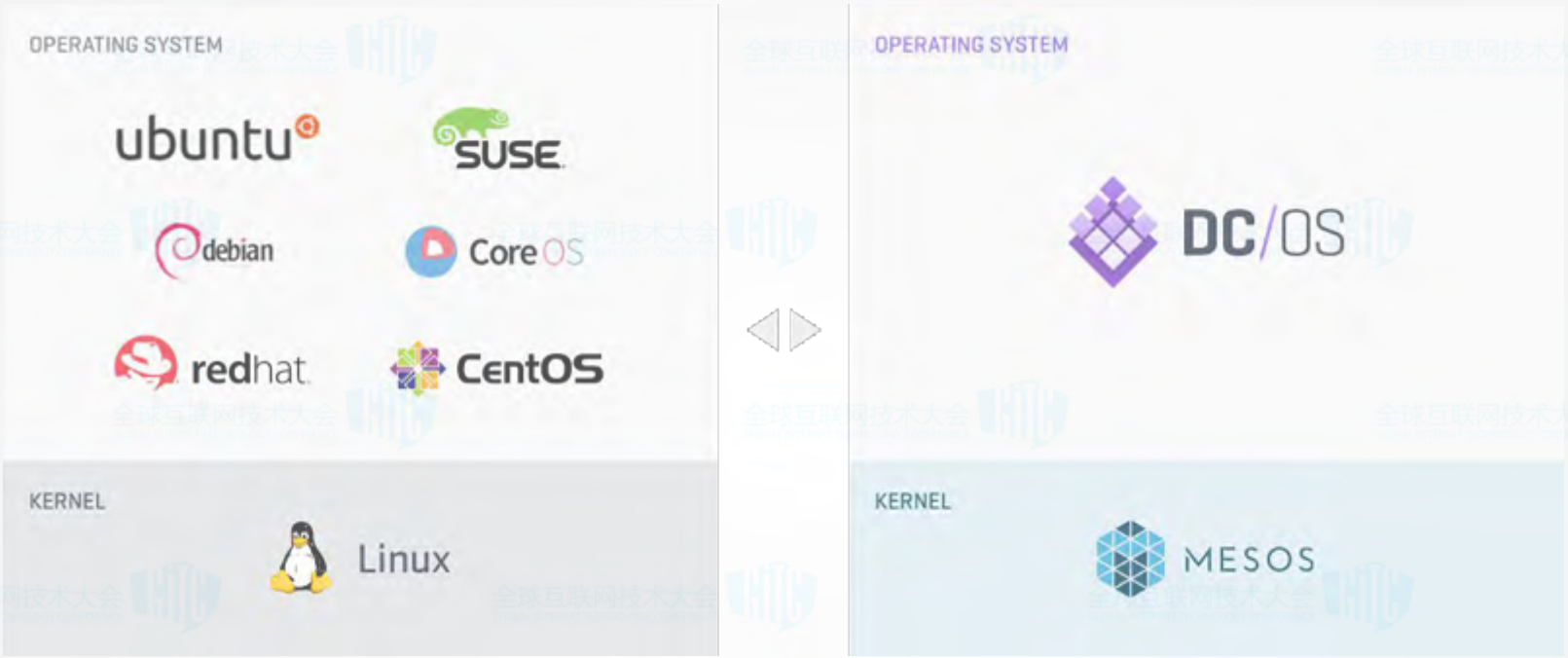
Jenkins

云上系统

mesos+Marathon+docker

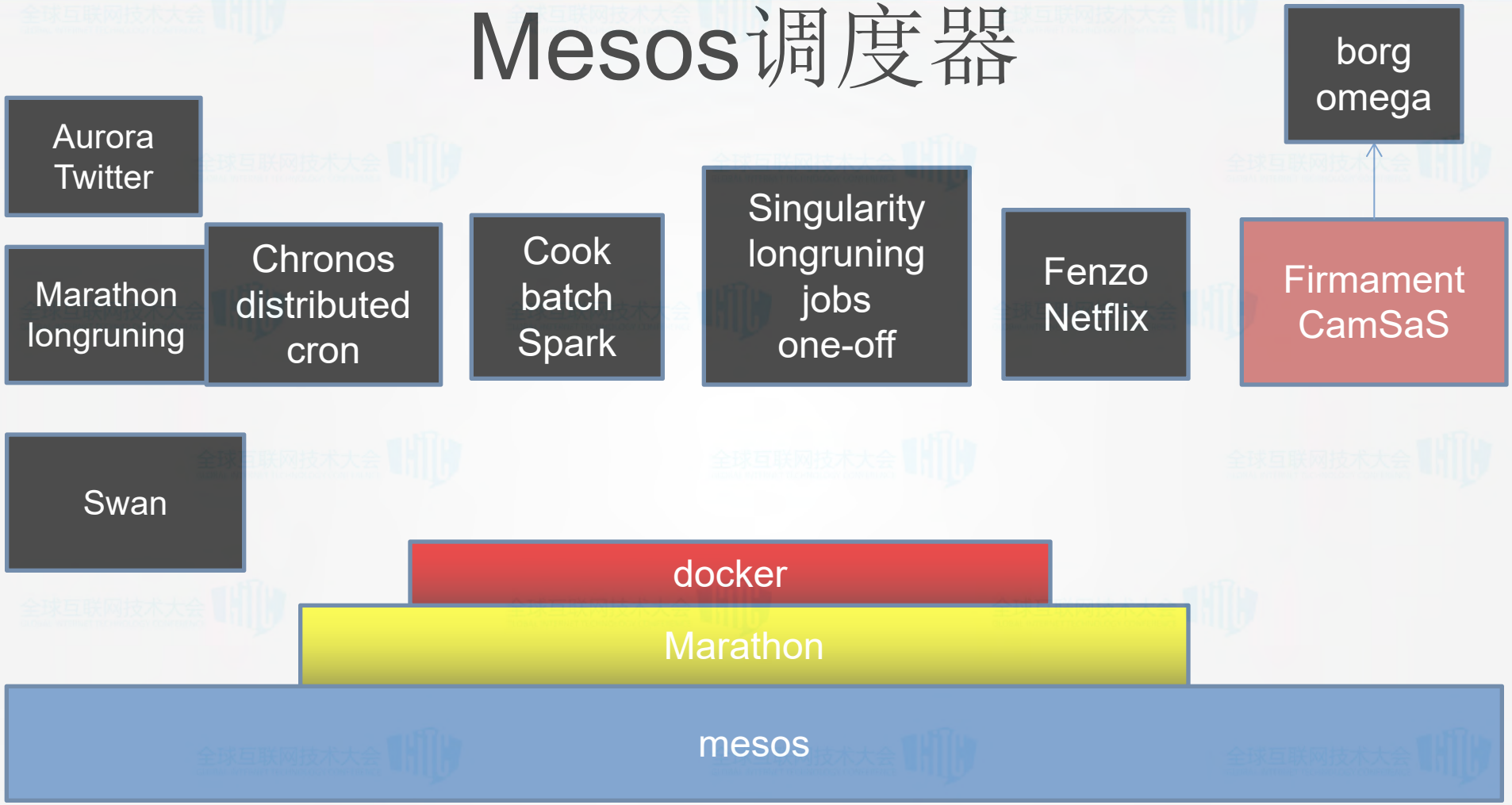
kubernetes+docker

dcos



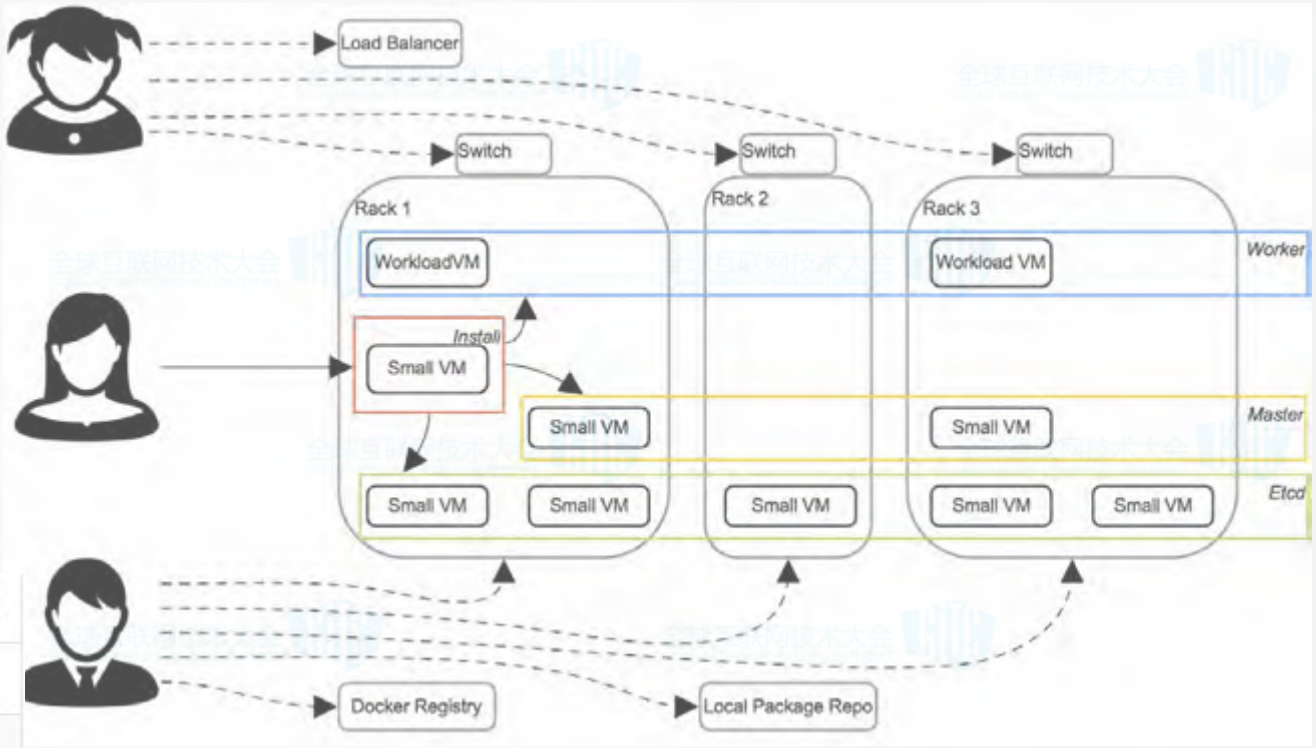
目前不支持docker1.12

Mesos调度器


borg
omega
Aurora
Twitter
Marathon
longruning
Chronos
distributed
cron
Cook
batch
Spark
Singularity
longruning
jobs
one-off
Fenzo
Netflix
Firmament
CamSaS
Swan
docker
Marathon
mesos

k8s集群自动化部署

Fully-Automated



Dependency	Current version
Kubernetes	1.4.5
Docker	1.11.2
Calico	1.6
Etcd (for Kubernetes)	3.0.13
Etcd (for Calico)	2.37

https://github.com/apprenda/kismatic

Tensorflow

深度学习

GPU

可扩展

模型并行

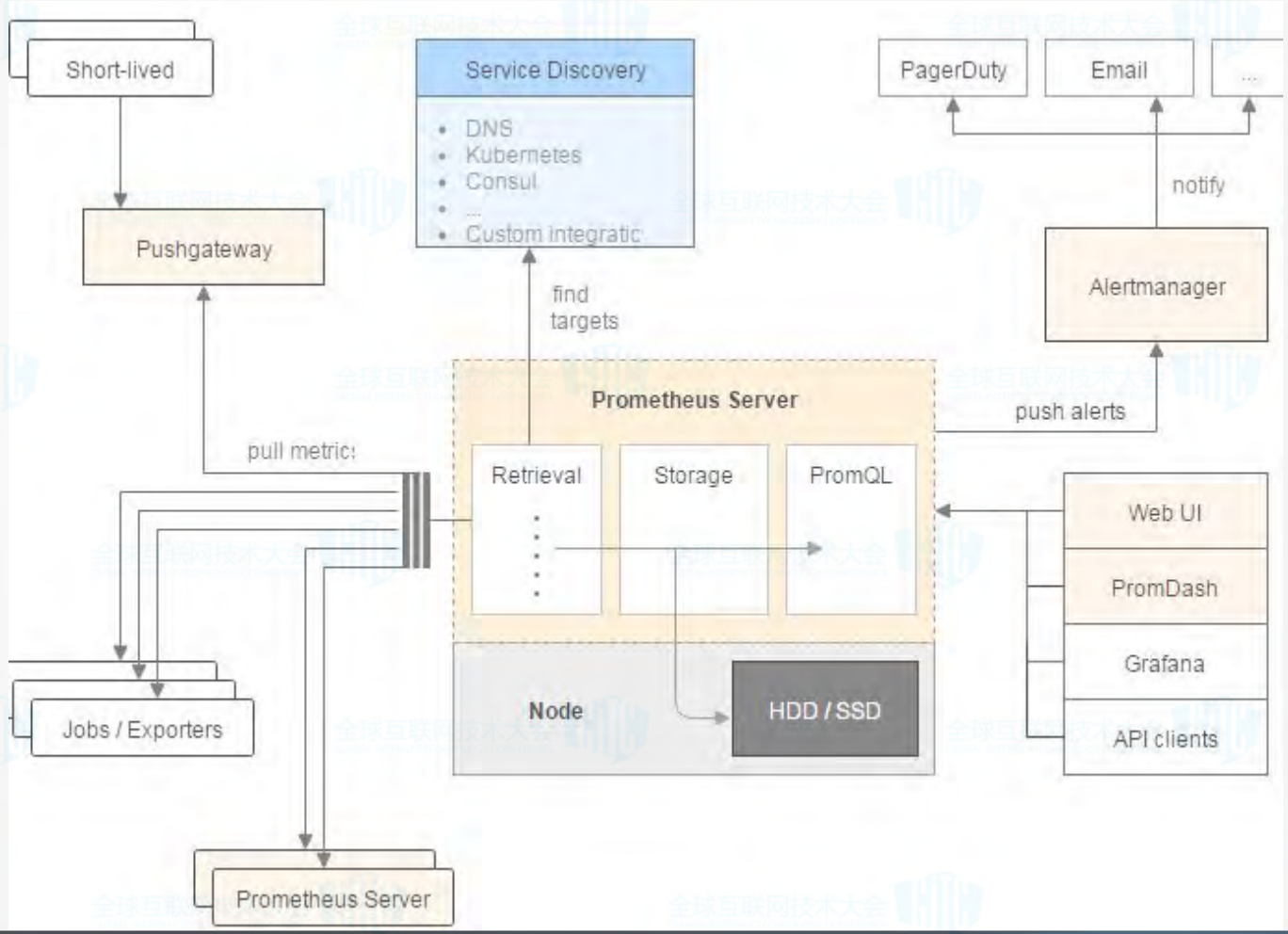
数据并行

https://github.com/douban/tfmesos

https://github.com/k8sp/k8s-tensorflow

Time Series metric DB

Prometheus push&pull

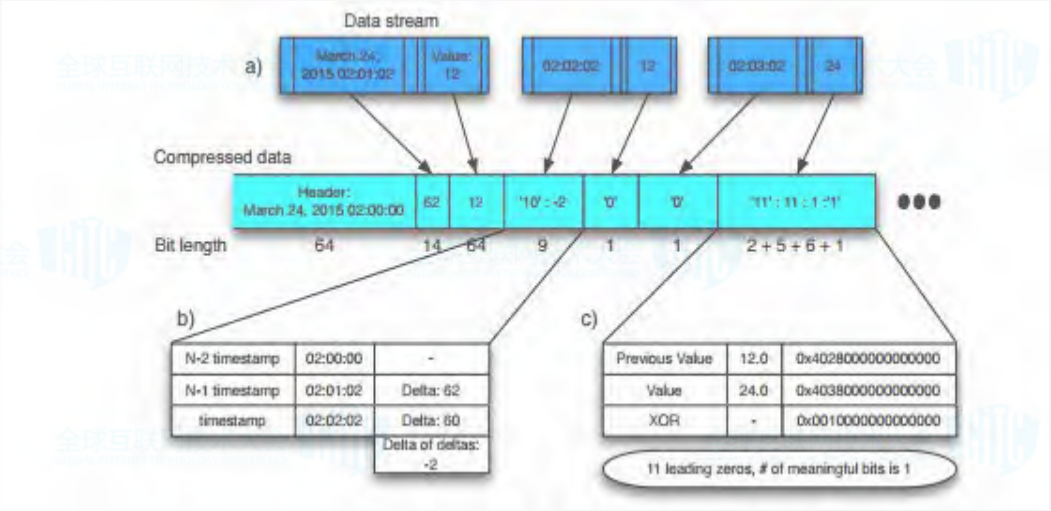


Prometheus

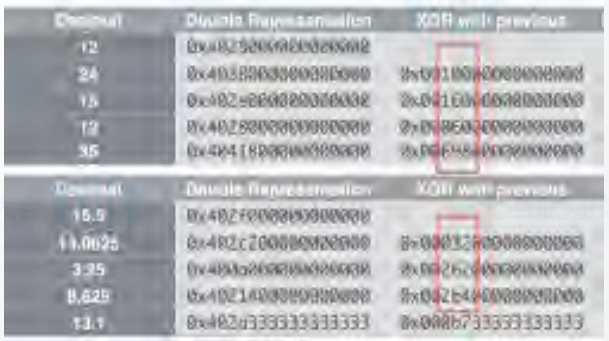
优点：块压缩算法

timestamp数据通过double-delta转bit

value采用xor异或control bit 做转化



https://github.com/prometheus/prometheus/blob/d93f73874f71288f26142f5264eae4585f14642b/storage/local/chunk/varbit.go



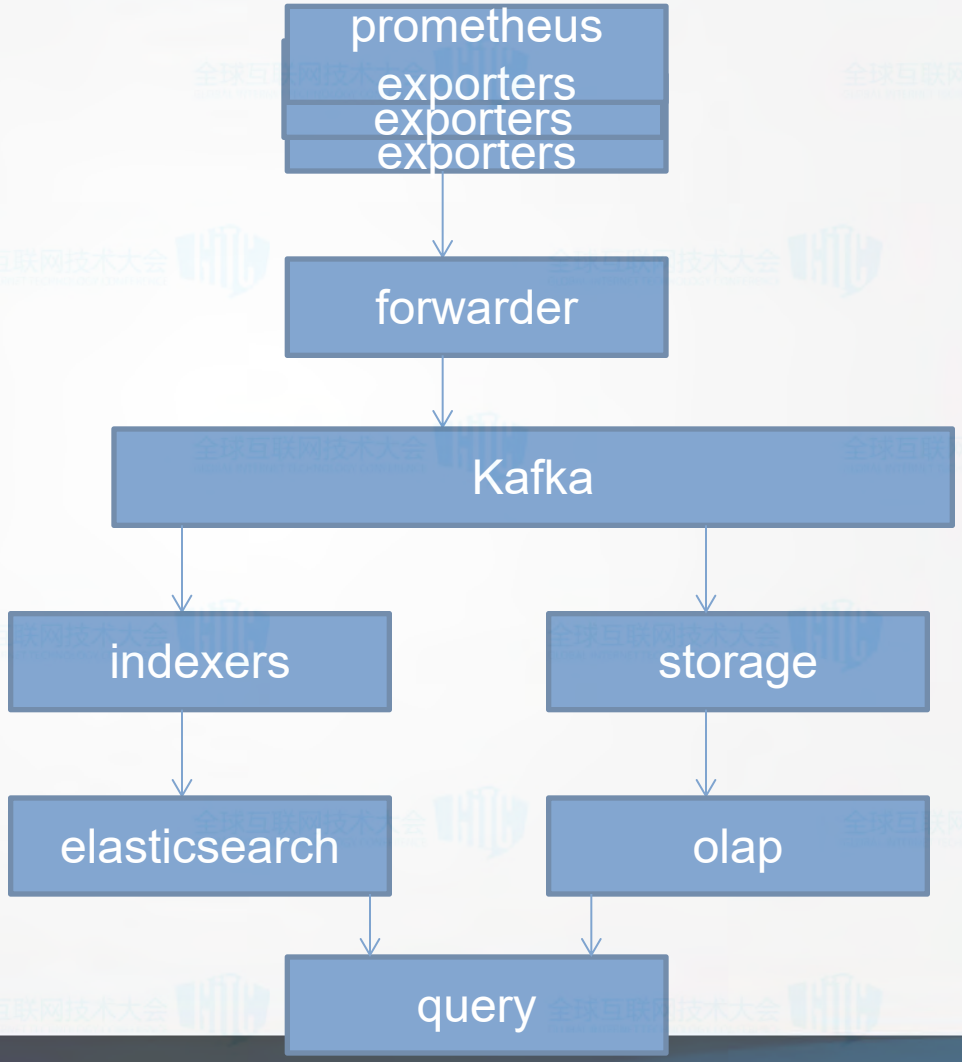
缺点：单机版

Prometheus分布式

1.分布式中间件（一致性hash）

2.pub-sub模式

https://github.com/digitalocean/vulcan



Auto Scaling

传统方法：基于规则模板
cpu util >80%
memory >80%
disk >80%

新方法：机器学习预测

分类，回归

卡尔曼滤波预测

服务发现

Zookeeper

Etcd

Consul

SmartStack airbnb

结论

节约机器资源

高效运维

DevOps趋势

机器解决机器问题是未来

谢谢观赏

联系方式：

微信：fengyuncrawl