

# 京东容器平台与数据中心协同发展实践

鲍永成

京东商城-基础平台部 技术总监

# SPEAKER INTRODUCE

---

## 鲍永成

- 2013年加入京东，负责京东容器集群平台（JDOS）研发，带领团队完成京东容器大规模落地战略项目，有效承载京东全部业务系统和中间件，以及数据库系统，特别在大促期间抢占式调度应对高峰流量，技术引领，从容大促。
- 目前聚焦在京东容器集群 JDOS 2.0 以及京东敏捷智能数据中心研发。



TABLE OF  
**CONTENTS 大纲**

---

- 容器技术在京东商城历史与未来
- 容器技术是什么
- 传统数据中心
- 京东敏捷数据中心建设
- 开源

# 问题

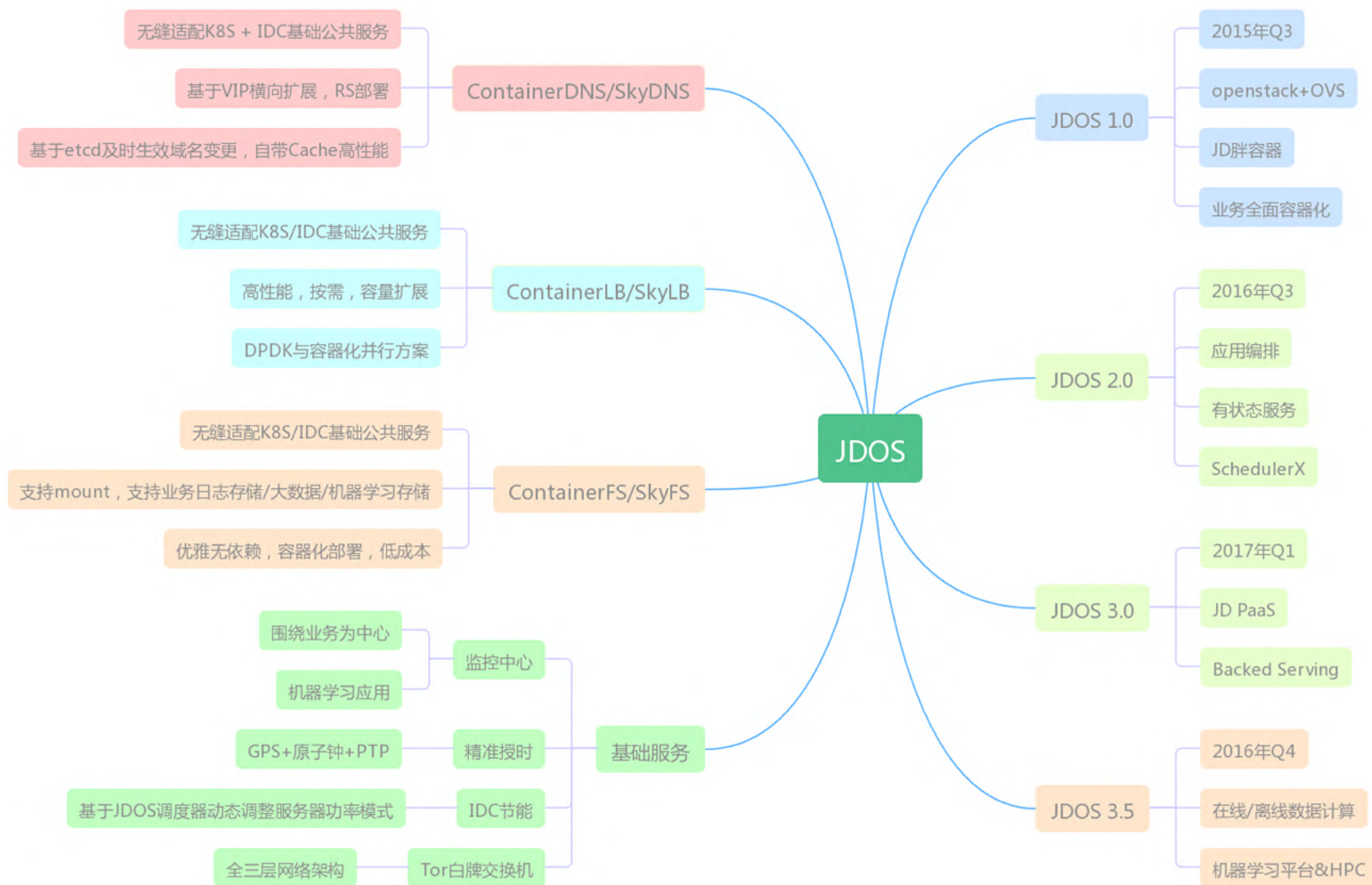
- 容器集群管理&应用编排 **VS** 数据中心基础软件

# 梦想

- 不止于分容器
- 让编排成为自然的事情

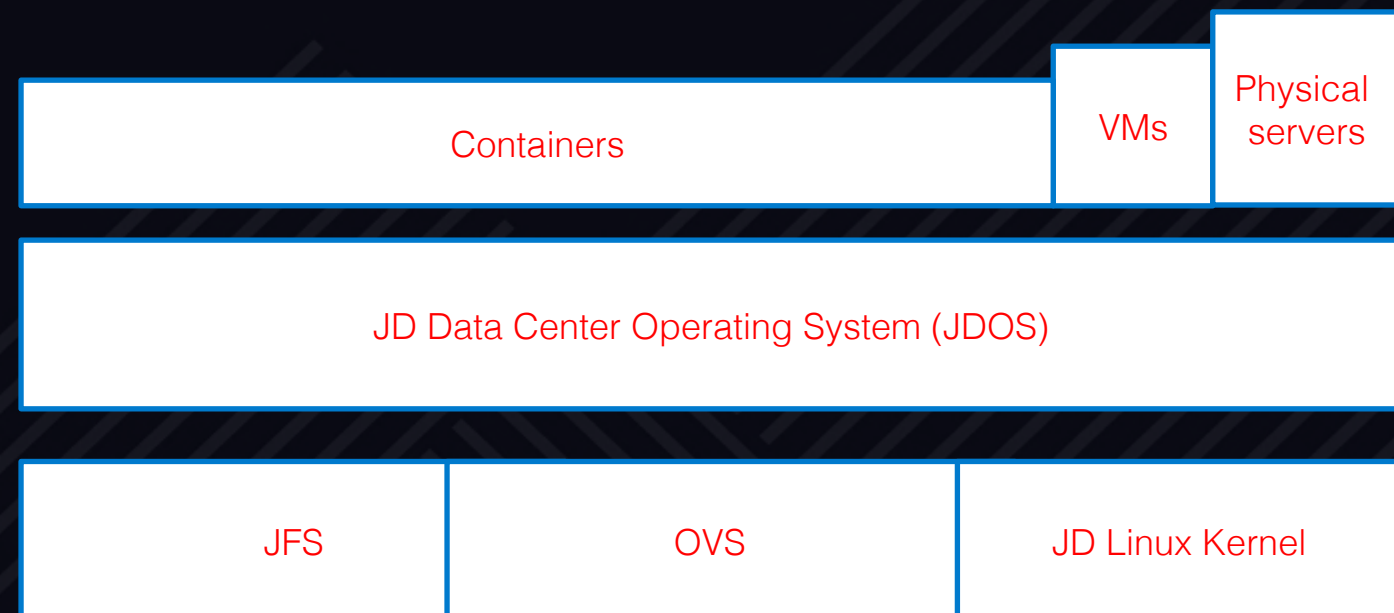


# JDOS 图谱



# JDOS 1.0

- 分物理机 => 分容器
- 业务系统100%全面容器运行
- 上线周期由1天降低到10分钟
- 少买3/4机器

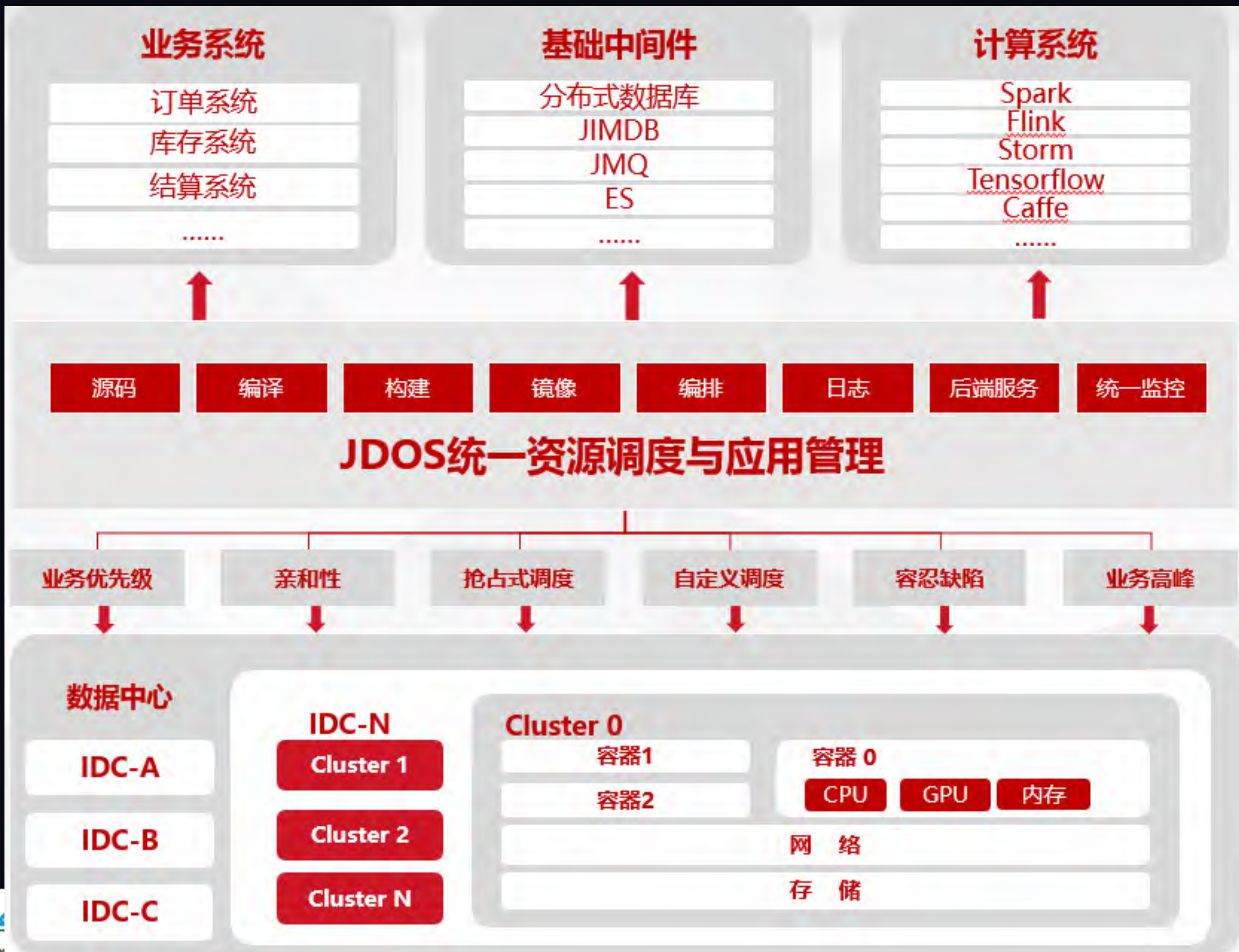


# JDOS 2.0 / 3.0 / 3.5

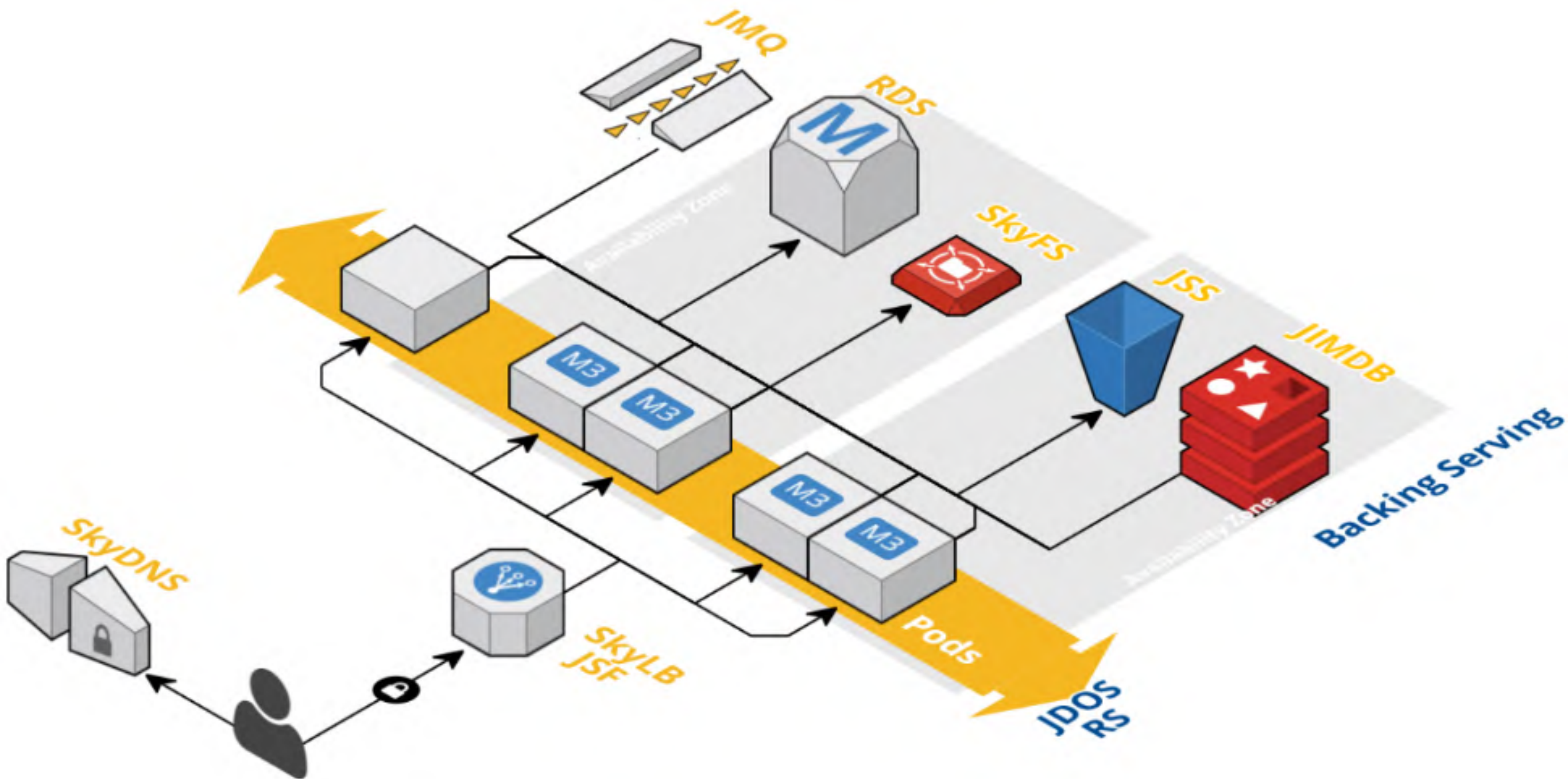
- 60%业务
- SRE初步形成
- 首次支撑618大促
- 全面基于镜像发布上线
- RS副本数 从容扩容应对高峰流量
- CPU利用率提升3倍，有效消除CPU/Network热点



# JDOS 2.0 / 3.0 / 3.5



# JDOS 2.0 / 3.0 运行典型业务

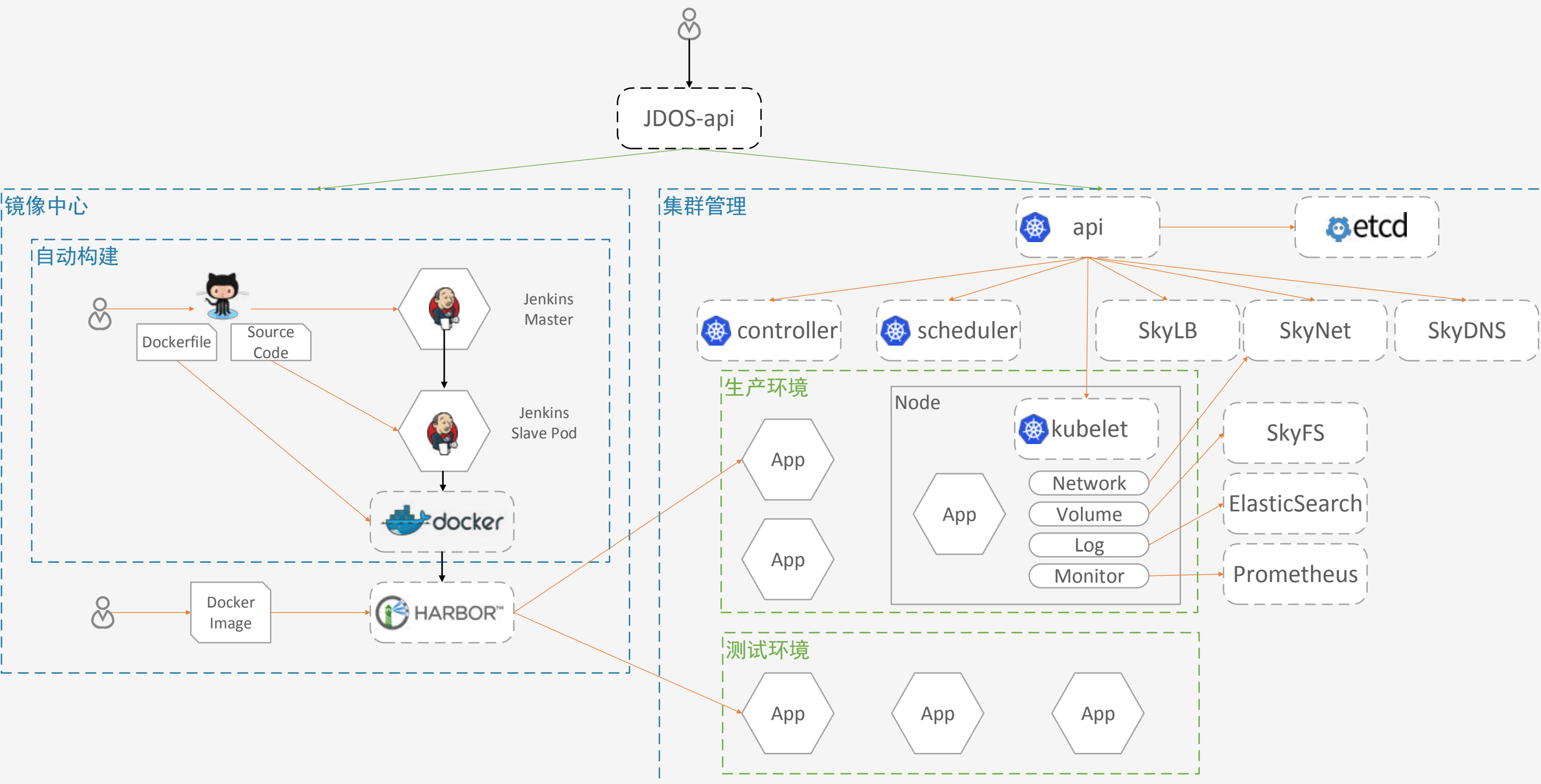


# JDOS 2.0 / 3.0 核心功能

- JD PaaS
- Self-Service , On-Demand Application Stacks
- CI , Build Image Services
- Logs , Monitors , APM Service
- DNS , LB , JSF Service
- Distributed Shared Storage Service
- Attached Resources Service
- App Orchestration



# JDOS 2.0 / 3.0 / 3.5 集群部署



## TABLE OF CONTENTS 大纲

---

- 容器技术在京东的历史和未来
- 容器技术【泛指Container&CM】是什么
- 传统数据中心
- 京东敏捷数据中心
- 开源



# 容器技术是什么-1/3

- Linux Kernel

# 容器技术是什么-2/3

- Networking

# 容器技术是什么-3/3

- Scheduler

Create

Resource Rebalance

## TABLE OF CONTENTS 大纲

---

- 容器技术在京东的历史和未来
- 容器技术是什么
- 传统数据中心
- 京东敏捷数据中心建设
- 开源计划

# 传统数据中心

- DNS
- 负载均衡
- 二层VLAN网络
- CMDB
- 7x24监控中心

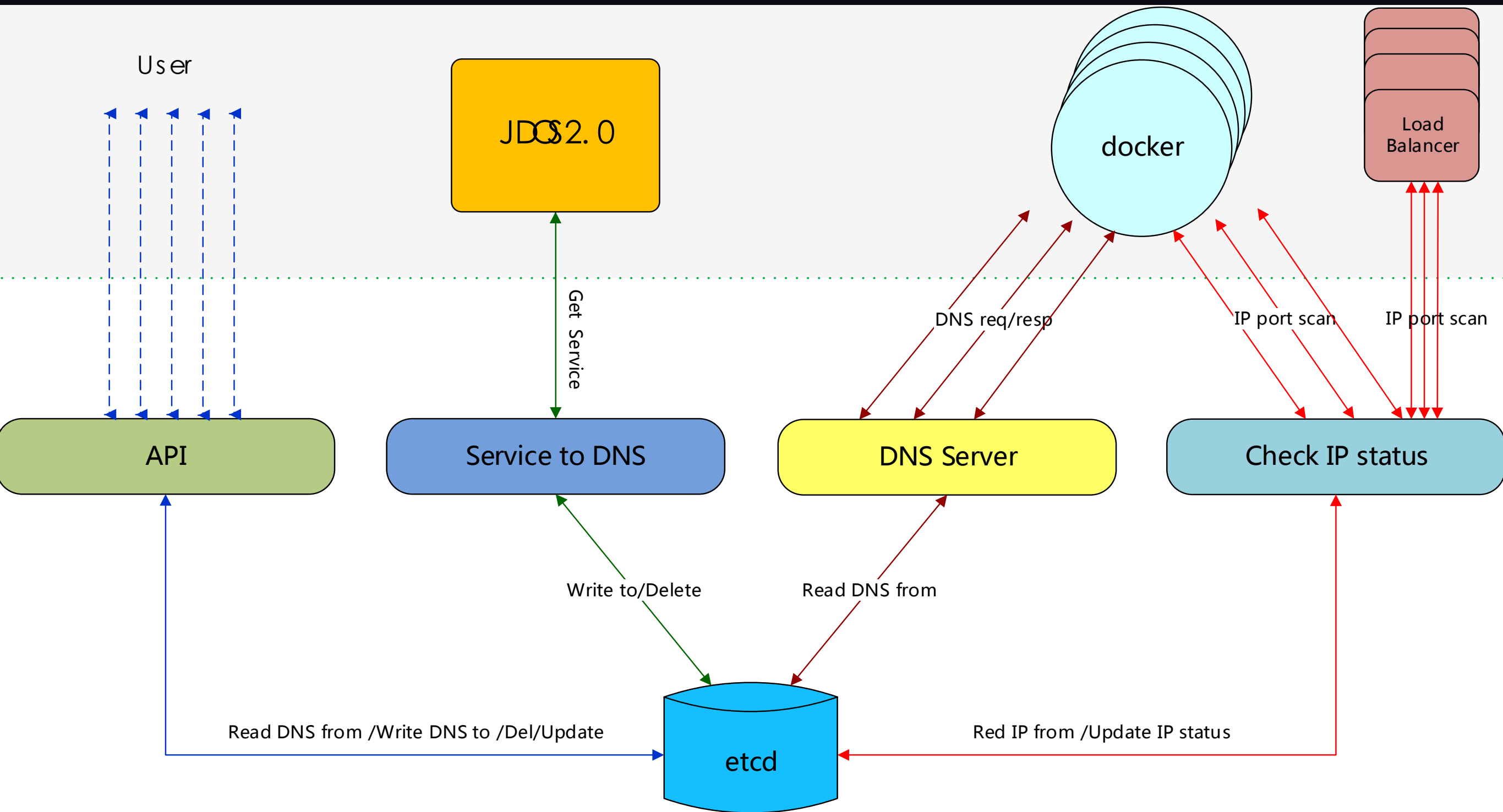


## TABLE OF CONTENTS 大纲

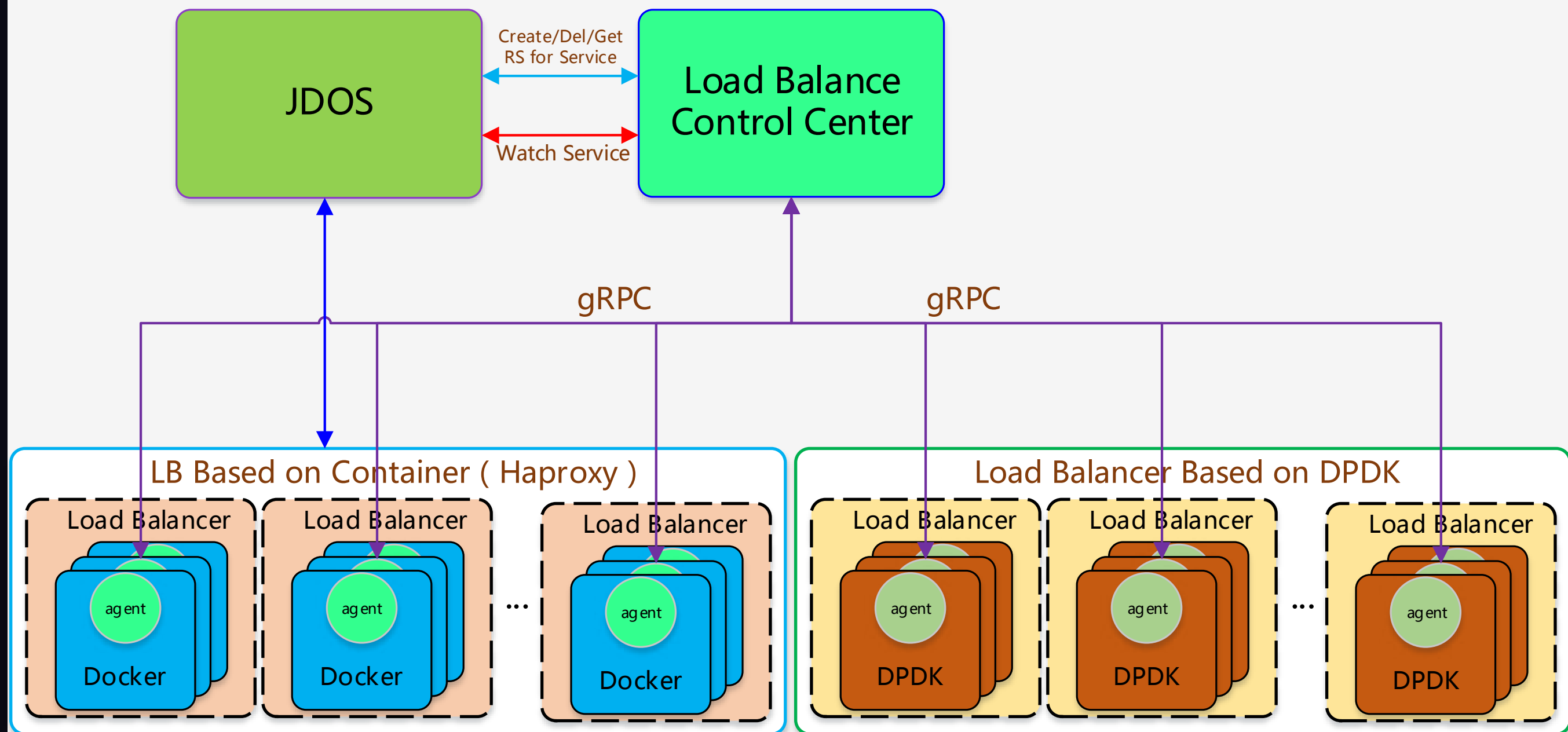
---

- 容器技术在京东的历史和未来
- 容器技术是什么
- 传统数据中心
- 京东敏捷数据中心建设
- 开源

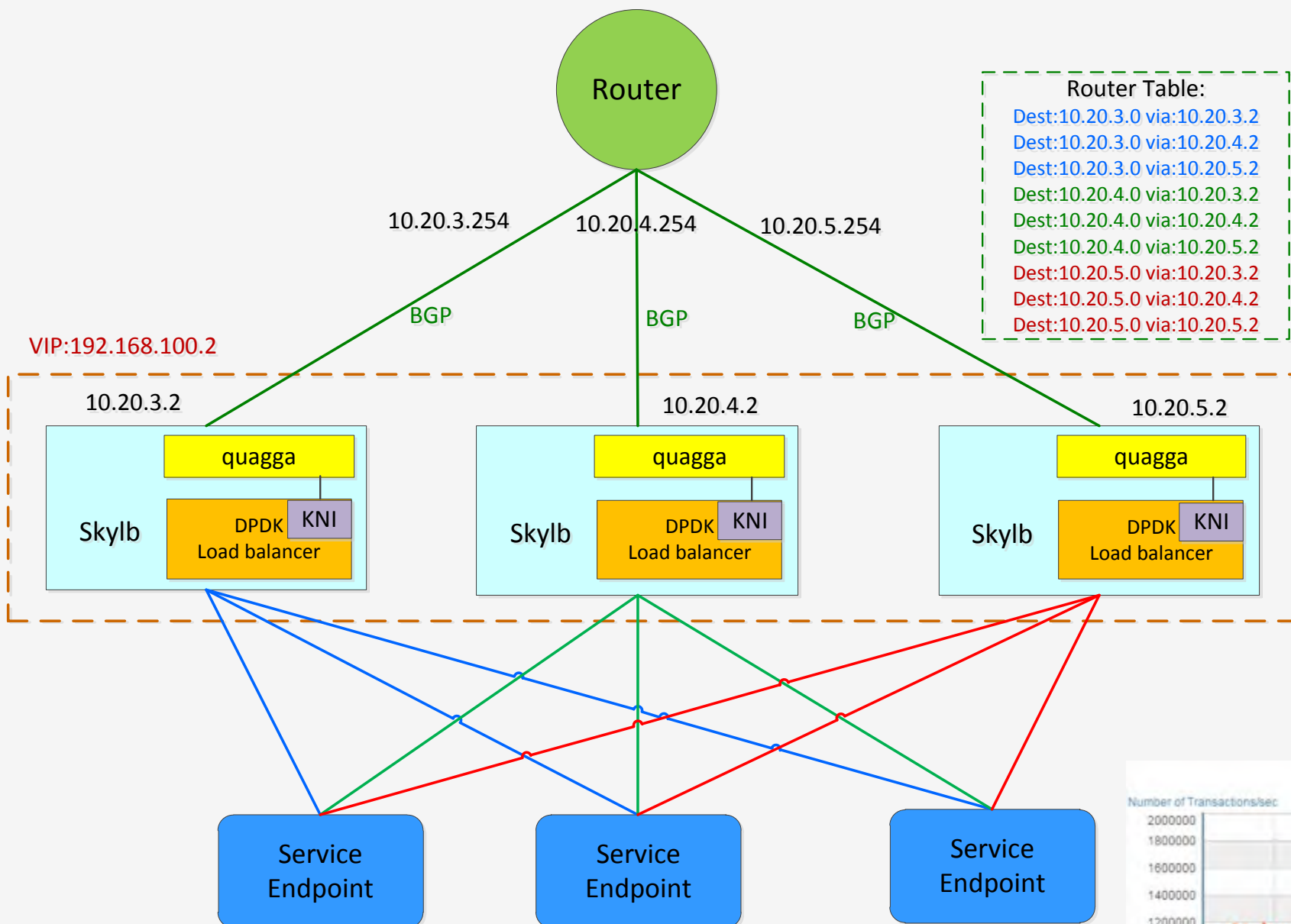
# 京东分布式 DNS



# 负载均衡服务部署架构



# 京东高性能 L4



Client-pkt-hashtable

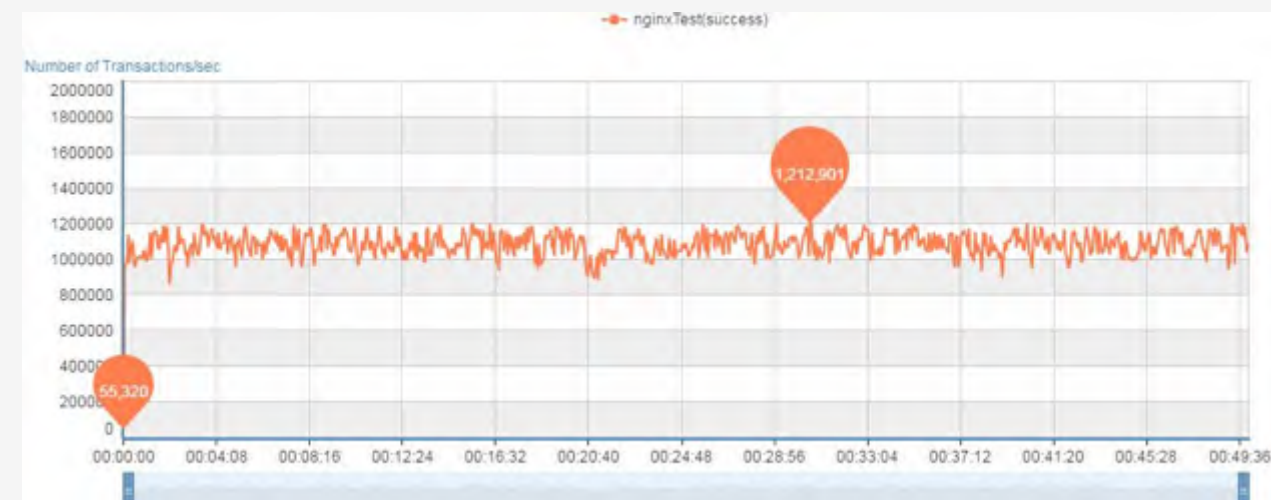
Hash 0
Hash 1
Hash 2
Hash 3
Hash 4
Hash 5
Hash 6
Hash 7

Session table

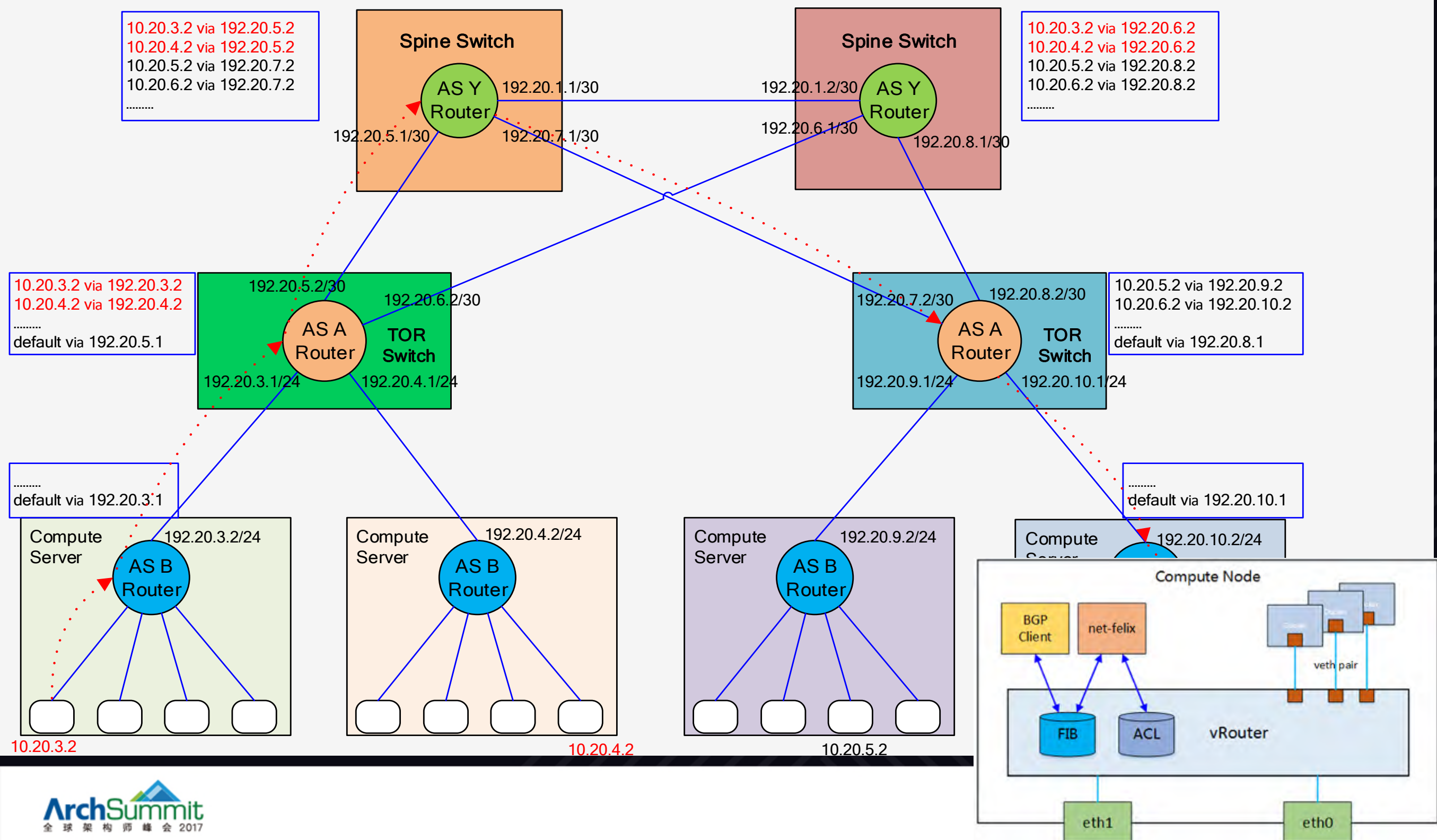
Session 0
Session 1
Session 2
Session 3
Session 4
Session 5
Session 6
Session 7

rs-pkt-hashtable

Hash 0
Hash 1
Hash 2
Hash 3
Hash 4
Hash 5
Hash 6
Hash 7

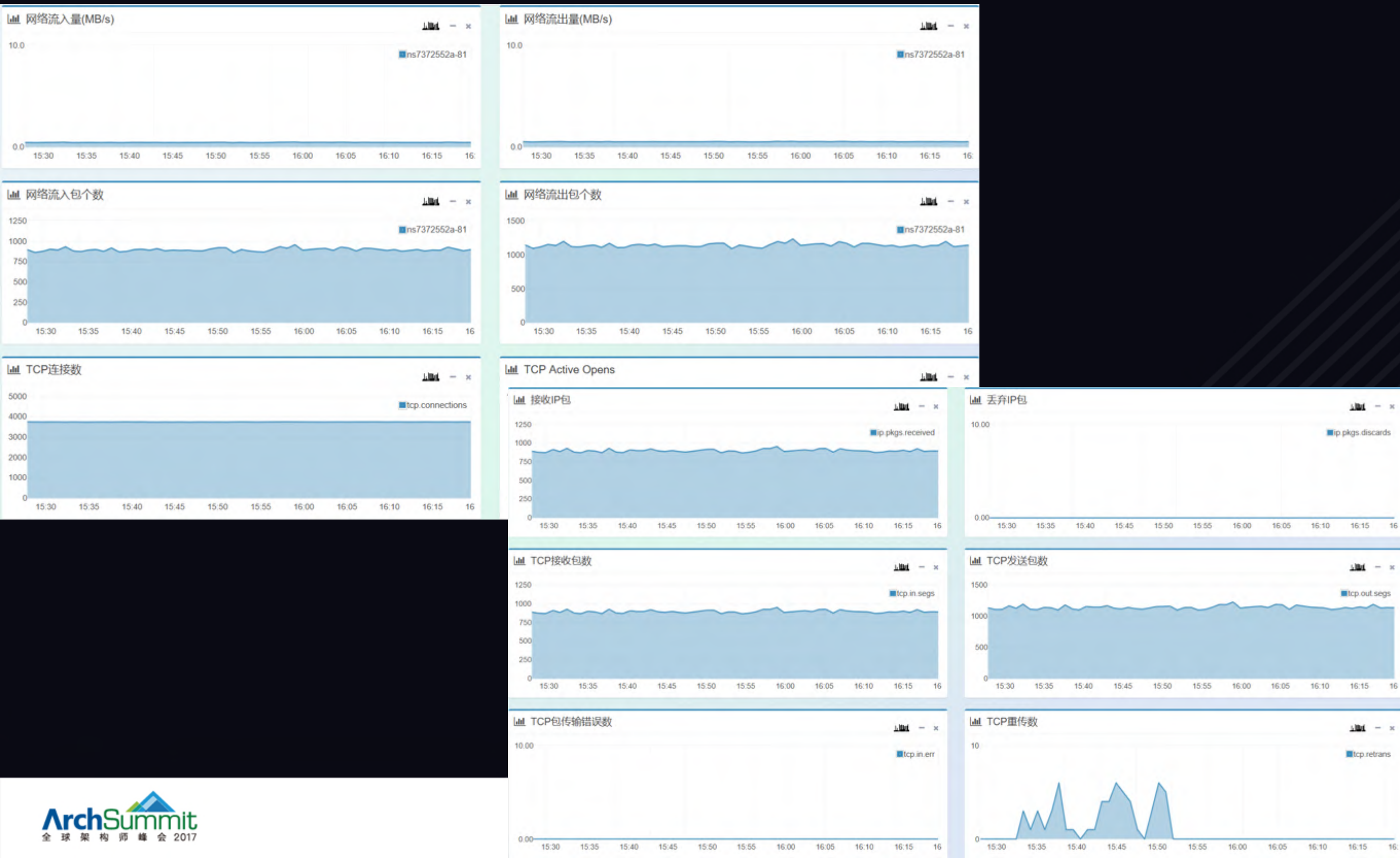


# 京东百万级容器网络

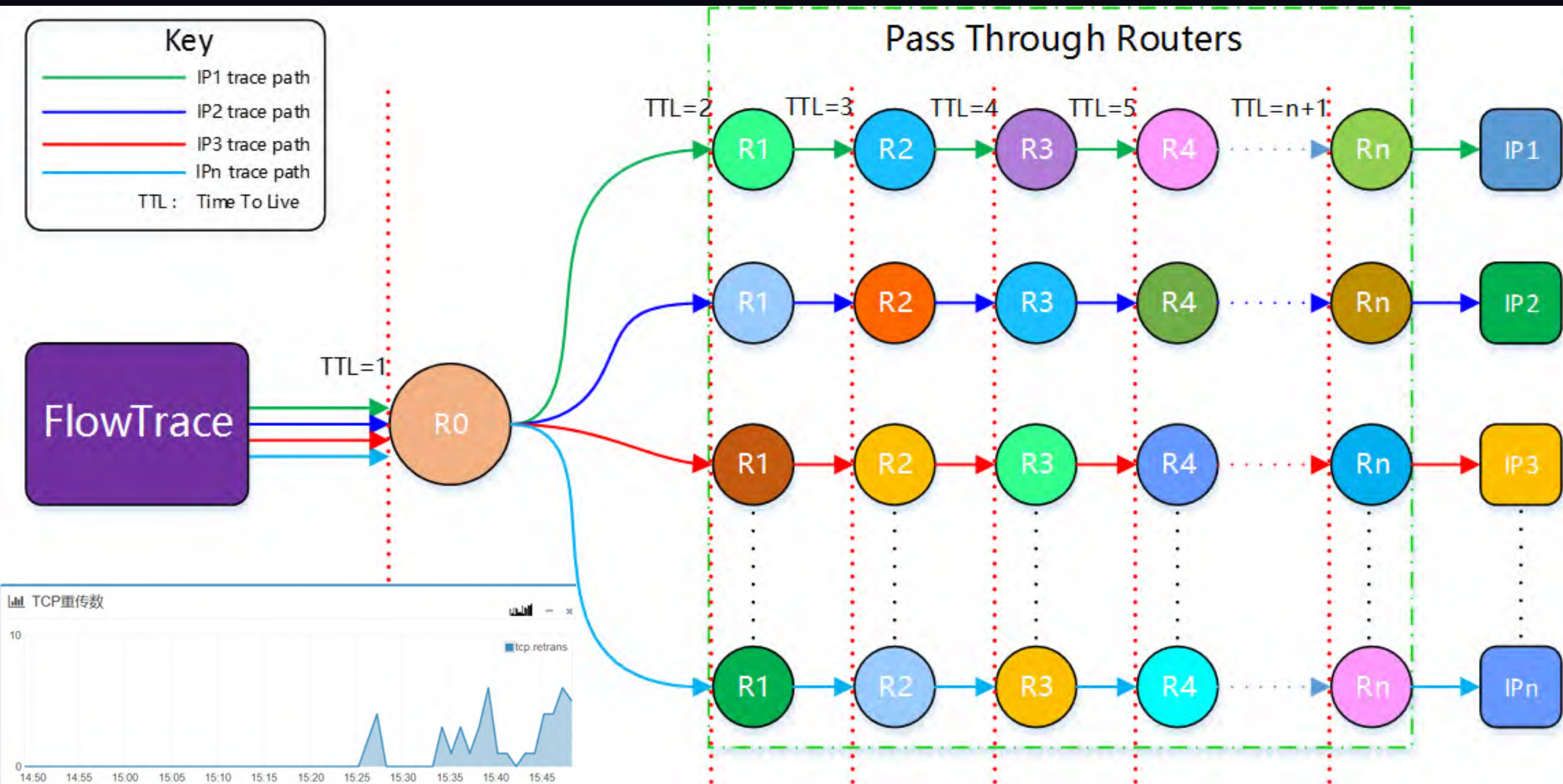




# 网络服务质量 1/2

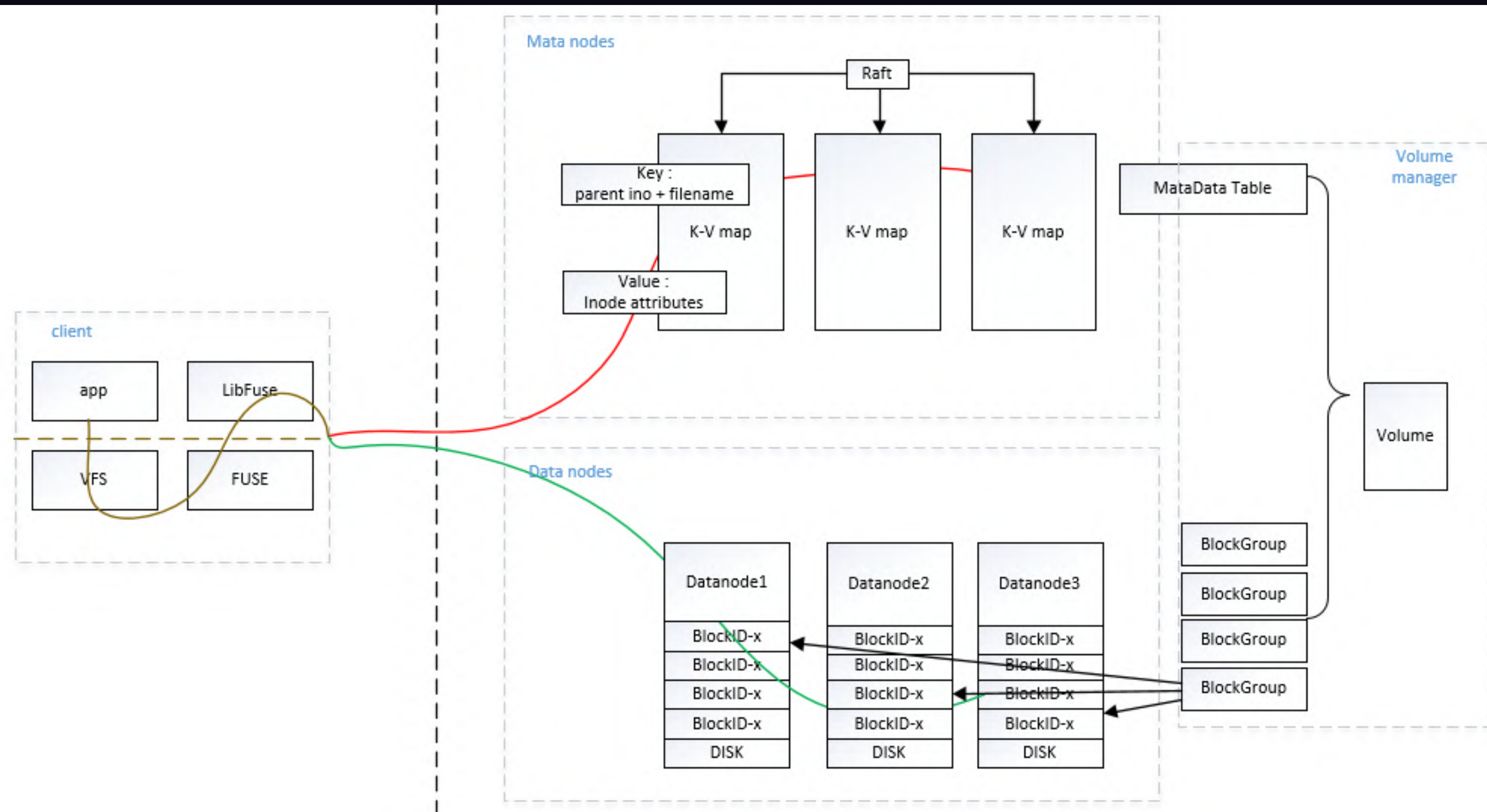


# 网络服务质量 2/2

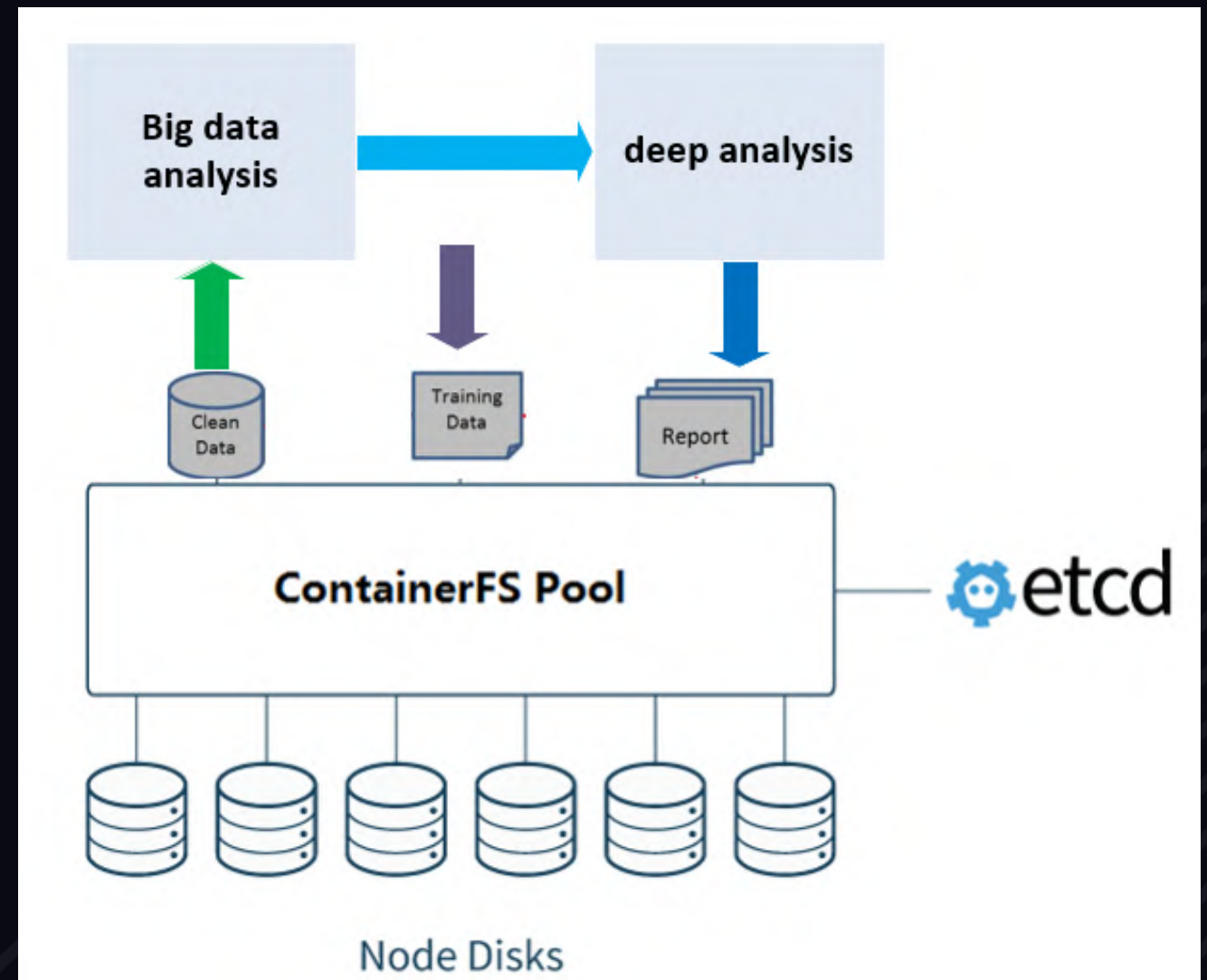
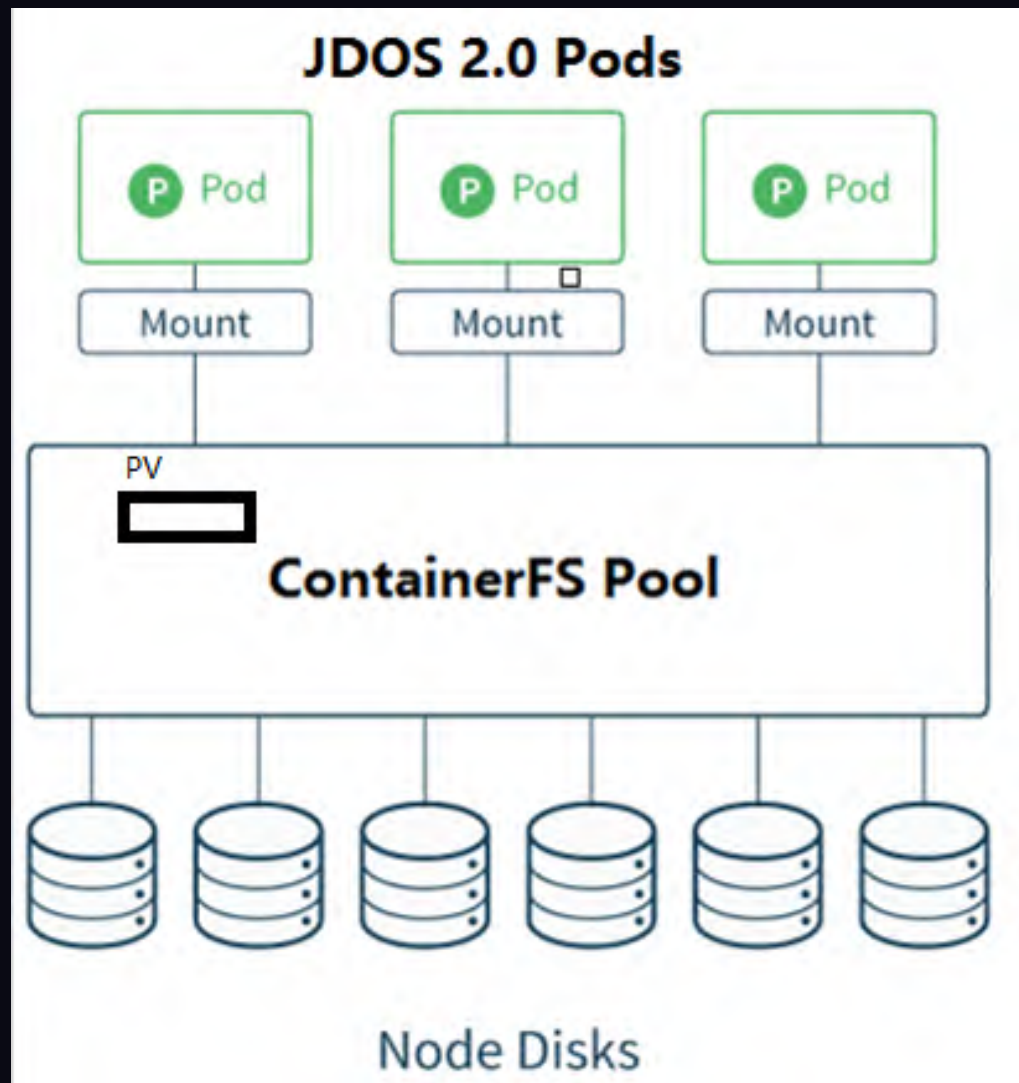




# 分布式共享存储-1/2



# 分布式共享存储-2/2



## TABLE OF CONTENTS 大纲

---

- 容器技术在京东的历史和未来
- 容器技术是什么
- 传统数据中心
- 京东敏捷数据中心建设
- 开源



# 开源计划

- 分布式共享存储

<https://github.com/ipdcode/containerfs>

- 分布式DNS

<https://github.com/ipdcode/skydns>

go report B build passing

## ContainerFS

a cluster filesystem for the containers

## Concepts

a volume = a metadata table + multiple block groups

## Key Problem: Ino Generation

auto increase on the client side, and fetch the initial ino from the metadata table on mount  
or allocate on the metadata table side

## Replication

raft or paxos for the metadata table

append-only write for blockgroups

## SkyDNS



## Introduction

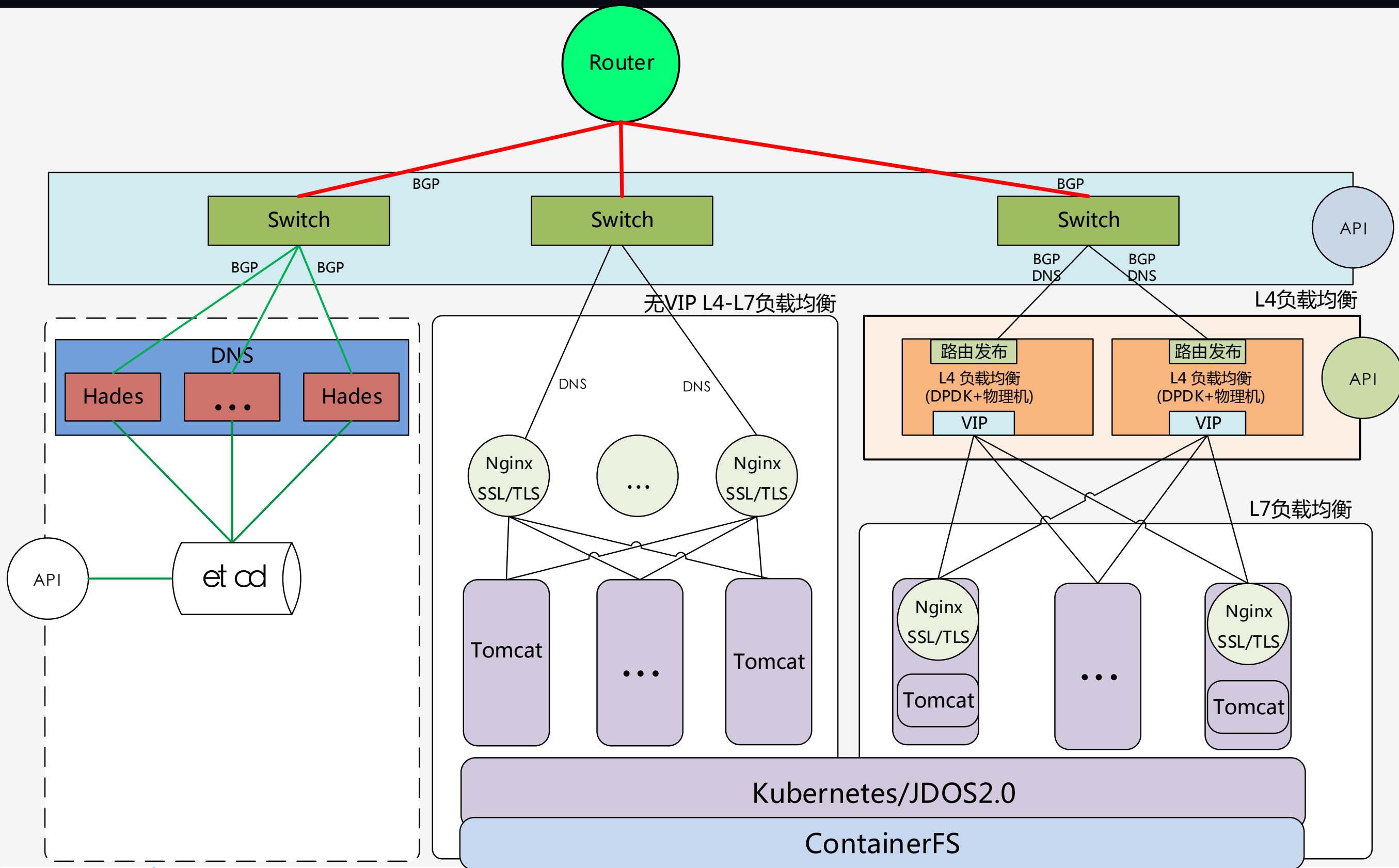
SkyDNS is used as internal dns server for k8s cluster, and use DNS library : <https://github.com/miekg/dns>. skydns-kubeapi will monitor the services in k8s cluster, when the service is created and has been assigned with external ips, the user(docker) in cluster can access the service with the domain. When the domain has multiple ips, the skydns will randomly choose one activated for the user, it seems like loadbalance. Also the skydns offer "session persistence", that means we query one domain from one user ip, then the user access the domain later, the user will get the same service ip.

## Components

- skydns : the main service to offer dns query.
- skydns-kubeapi : monitor the changes of k8s services, and record the change in the etcd. It offered the original data for skydns, meanwhile skydns-kubeapi offers the restful api for users to maintain domain records.
- skydns-apicmd : it is a shell cmd for user to query/update domain record, it is based on skydns-kubeapi.

# THANKS!

# JD IDC 容器&基础软件



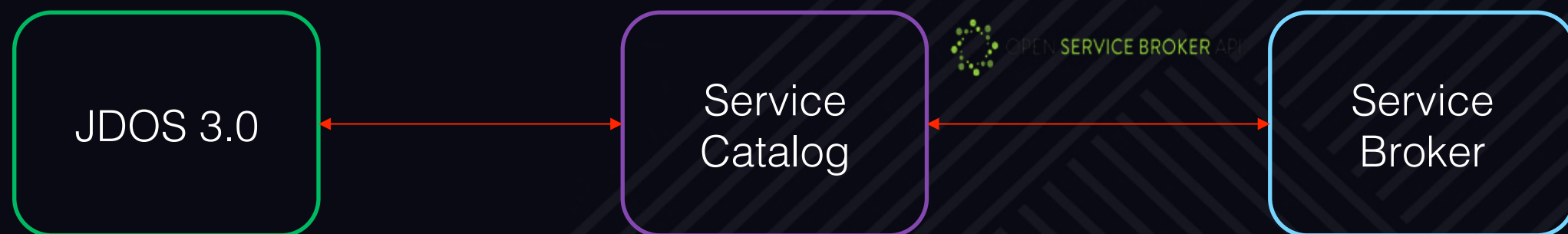
# JDOS 3.0 Backing Services 集成方式

## □ JDOS3.0 Service Catalog

通过标准的 [Open Service Broker API](#) 集成各个后端服务

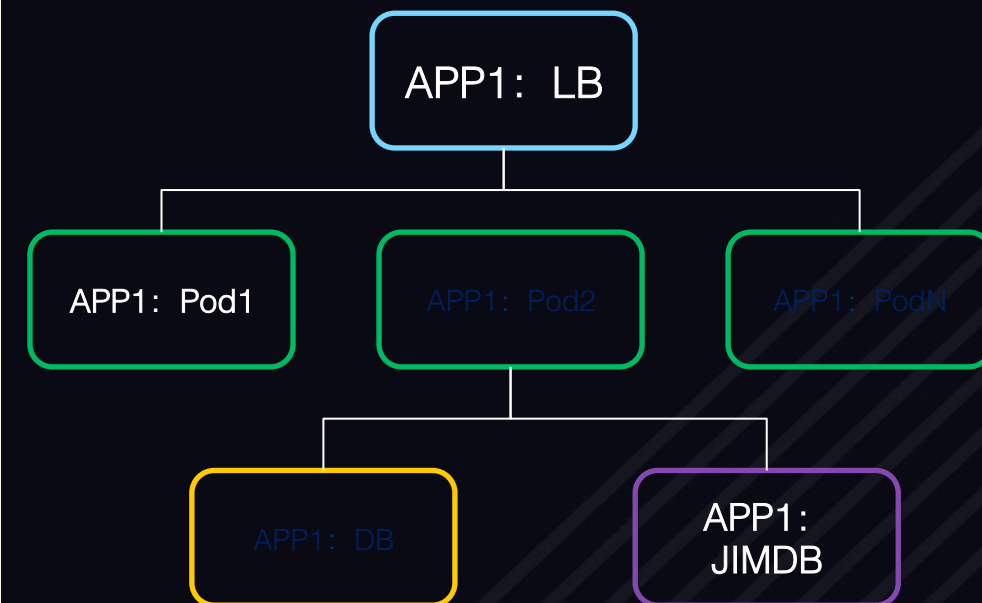
## □ Service Broker

实现了 [Open Service Broker API](#) 的一组后端服务



# JDOS 3.0 Backing Services 模版

```
apiVersion: jdoss.jd.com/v1beta1
kind: ServiceBroker
metadata:
  labels:
    app: appid1
  name: jimdb-appid1
spec:
  name: jimdb-appid1
  version: 2.8.3
  username: xxxxxx
  password: xxxxxx
  url: xxx-servicebroker.app.n.jd.local
```



🏠 / 🏠 分组管理

master

worker

zeppelin

+ 新增

机房

🏠

环境

生产集群

删除当前分组

配置管理

集群管理

日志服务

操作

### 基本配置

必选

编辑

容器规格	part(8C16384Mi)	镜像	public/spark:2.1.1_v3
分组标签1	tag0	分组标签2	tag1
分组标签3	tag2	是否自动拉起	否
磁盘规格(GB)	100	磁盘类型	SAS
CPU型号		zone	

### 高级配置

可选

编辑

容器变量

变量名	值
SPARK_DAEMON_MEMORY	4G
SPARK_LOCAL_DIRS	/export/spark/data
NODE_TYPE	master
UI_PORT	8080

配置文件

文件及文件路径	是否包含加密内容	
/etc/jdos_host	否	
/etc/logrotate.d/jdos/logrotate	否	

📁 镜像管理

👤 分组管理

⚖️ 负载均衡



