

# Pouch和阿里容器技术演进

阿里巴巴系统软件事业部-林轩



# QCon

全球软件开发大会

## 成为软件技术专家 的必经之路

[北京站] 2018

2018年4月20-22日 北京·国际会议中心

**7折** 购票中, 每张立减2040元  
团购享受更多优惠



识别二维码了解更多



# 极客时间

重拾极客精神·提升技术认知

## 下载极客时间App

获取有声IT新闻、技术产品专栏，每日更新



扫一扫下载极客时间App

# AiCon

全球人工智能与机器学习技术大会

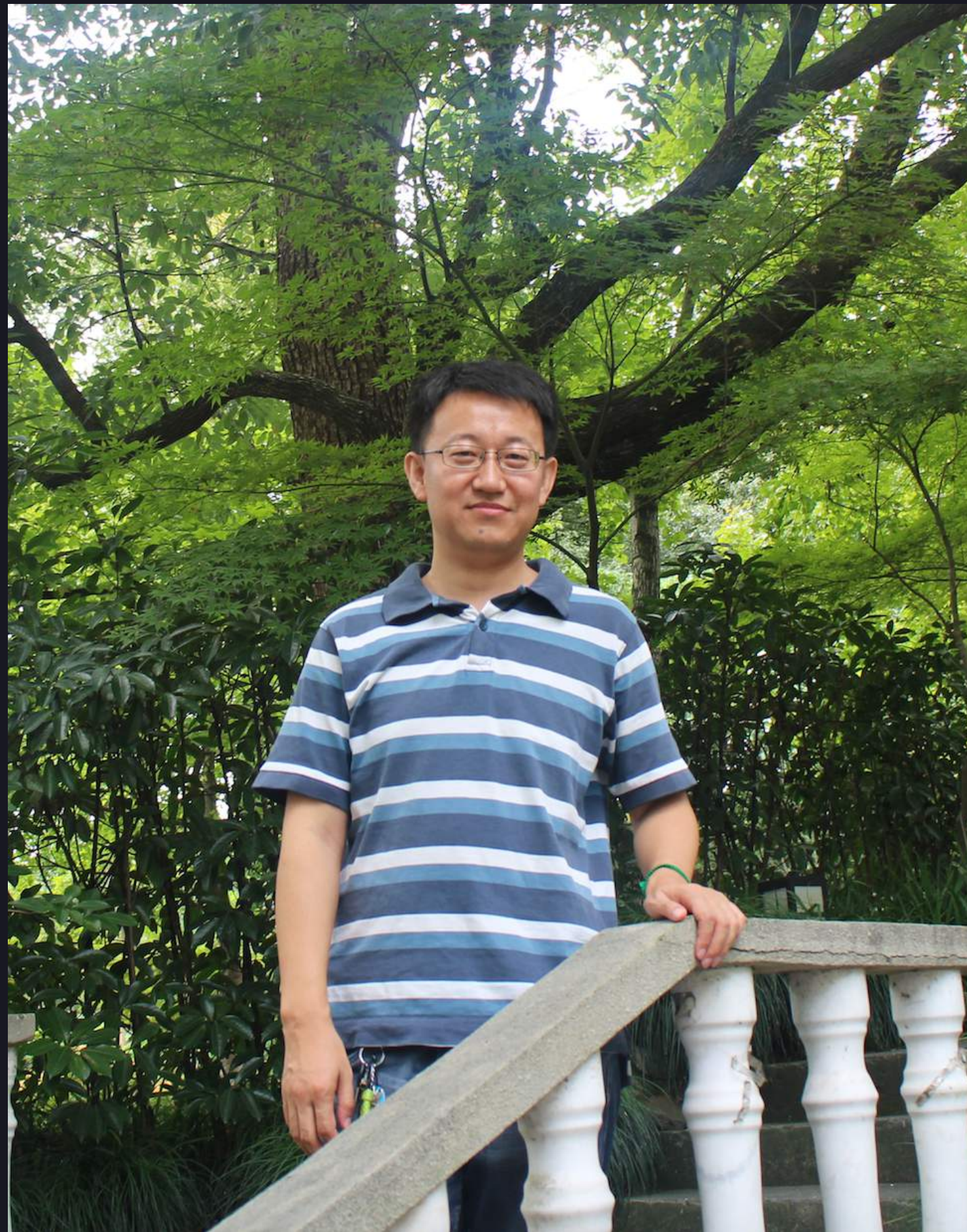
助力人工智能落地

2018.1.13 - 1.14 北京国际会议中心



扫描关注大会官网

# SPEAKER INTRODUCE



## 侯前明

阿里巴巴资深技术专家

花名林轩，2009年加入阿里巴巴，负责中间件分布式数据层TDDL

2011年开始孵化TAE项目，建设了TAE第一代和第二代平台

2015年开始转向基础设施建设，负责电商和大数据的混布实施。先后打造了Pouch容器和Sigma调度引擎，目前专注于容器技术在集团内外的推广和迭代，以及不同形态业务负载的混部和调度

TABLE OF

**CONTENTS** 大纲



**Pouch在阿里内部的使用现状**  
阿里容器技术的演进和实践  
Pouch的技术特点和开源路线

# Pouch在阿里内部的使用现状

## Pouch的规模：

- 服务了阿里集团和蚂蚁金服大部分BU
- 2017年双11容器百万级别
- 活跃应用100%完成了容器化

## 使用了Pouch的场景：

- 运行模式- 标准APP，购物车，广告、测试环境
- 编程语言- Java，C/C++，Nodejs，Golang
- 技术栈- 电商，DB，流计算，大数据，专有云

## 使用了Pouch的业务：

- 蚂蚁&交易&中间件
- B2B/CBU/ICBU/1688/村淘
- 搜索&广告&数据库
- 合一集团(优酷) &高德&UC
- 菜鸟&钉钉&盒马
- 安全&无线&阿里云专有云输出
- 集团测试环境
- .....

TABLE OF

**CONTENTS** 大纲



Pouch在阿里内部的使用现状  
阿里容器技术的演进和实践  
Pouch的技术特点和开源路线



# Pouch 阿里从物理机到容器的演变

## 架构演变

- 运行：从集中式到分布式
- 运维：从分散到云化

## 资源使用的演变

- 从物理机到VM
- 从VM到容器

## 容器的要素--阿里内部运维和应用视角

- 有独立IP
- 能够ssh登陆
- 有独立的，隔离的文件系统
- 资源隔离—使用量和可见性

和物理机的  
使用体验一致

# Pouch 阿里容器发展轨迹

## 容器的要素--阿里内部运维和应用视角

- 有独立IP
- 能够ssh登陆
- 有独立的，隔离的文件系统
- 资源隔离—使用量和可见性

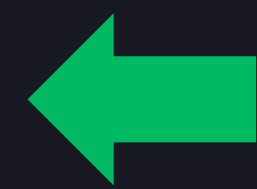
## 手工Hack实现容器要素

- ➡ 虚拟网卡，网桥
- ➡ sshd
- ➡ Chroot (pivot\_root)
- ➡ CGroup，Namespace



引入LXC (Linux Container)  
内核可见性隔离Patch  
内核磁盘空间配额Patch

 Pouch



T4

引入Docker镜像

# Pouch 软件的本质

软件 = 文件(集) + 进程(组)

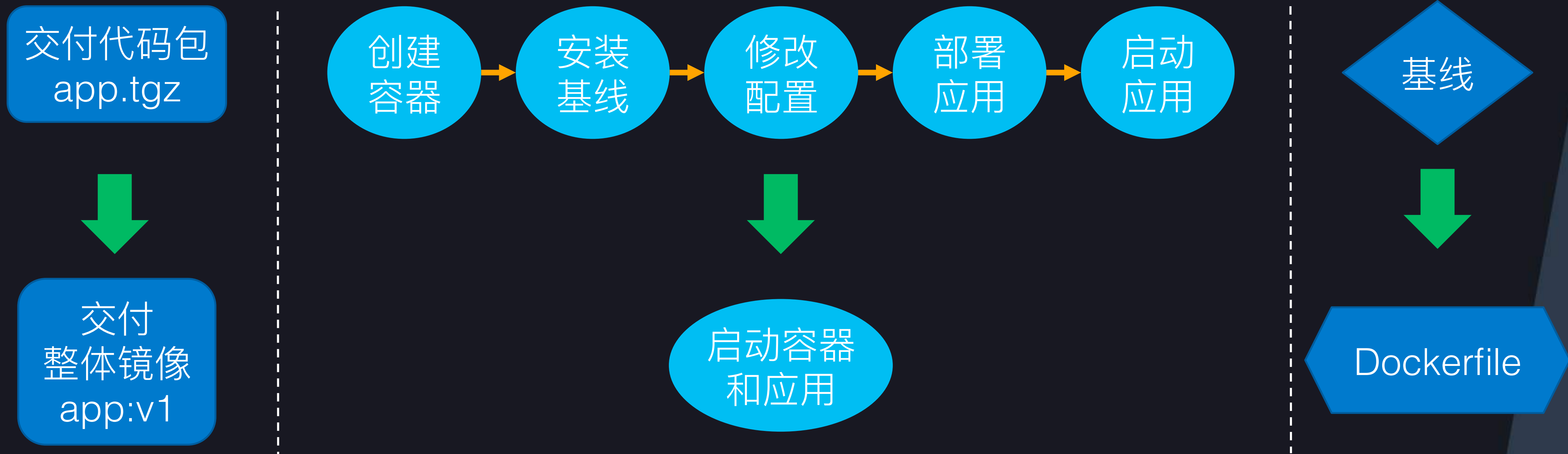
静态来看：文件 -- 一个有依赖层次的文件集  
动态来看：进程 -- 一个有交互关系的进程组



# Pouch T4和Docker的整合

2015年	T4	Docker
cpu隔离	cgroup.cpuset,cpu,cpuacct	cgroup.cpuset,cpu
内存隔离	cgroup.memory	cgroup.memory
进程隔离	pid namespace	pid namespace
文件系统隔离	chroot	chroot
磁盘空间隔离	dirquota	无
网络带宽隔离	飞天的netqos模块	无
网络模式	网桥+独立IP	网桥、Host
镜像系统	只有简单的t4模板	完备的Docker镜像系统
执行引擎	LXC	LXC , libcontainer
文件系统分层	overlayfs	aufs , device mapper , overlayfs
资源可见性	内核补丁隔离	无隔离
支持的内核和OS	alios5u,6u; 2.6.32-358; ali1172	7u+, 3.10+

# Pouch 研发和运维的改变



# Pouch 研发关注的要素

## 传统要素

- 功能与性能
- 稳定性
- 可扩展性
- 可测试性

## DevOps要素

- 可运维性
- 运维成本



DevOps八荣八耻

以可配置为荣，以硬编码为耻  
以系统互备为荣，以系统单点为耻  
以随时可重启为荣，以不能迁移为耻  
以整体交付为荣，以部分交付为耻  
以无状态为荣，以有状态为耻  
以标准化为荣，以特殊化为耻  
以自动化工具为荣，以人肉操作为耻  
以无人值守为荣，以人工介入为耻

# Pouch 阿里容器Roadmap

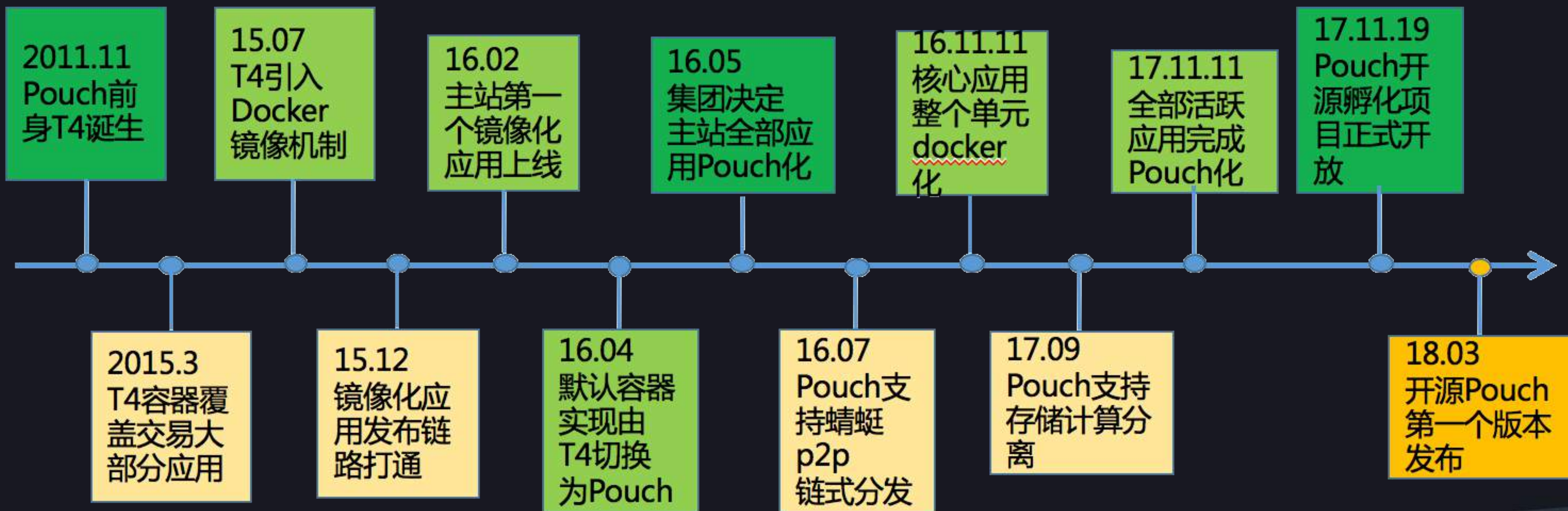


TABLE OF

**CONTENTS** 大纲



Pouch在阿里内部的使用现状  
阿里容器技术的演进和实践  
**Pouch的技术特点和开源路线**



# Pouch 技术优势



隔离性



P2P镜像分发



富容器



规模化考验

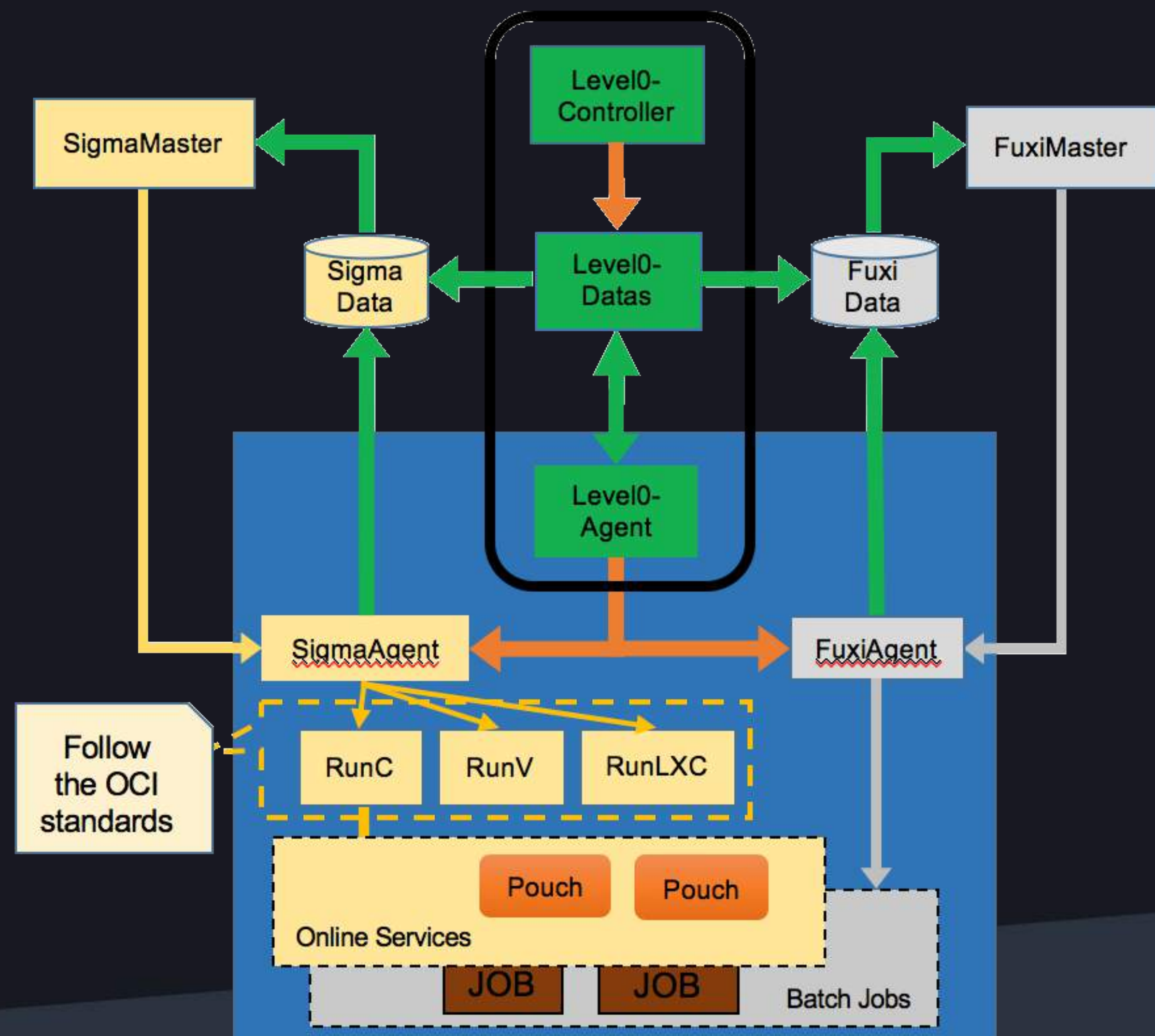


内核兼容性

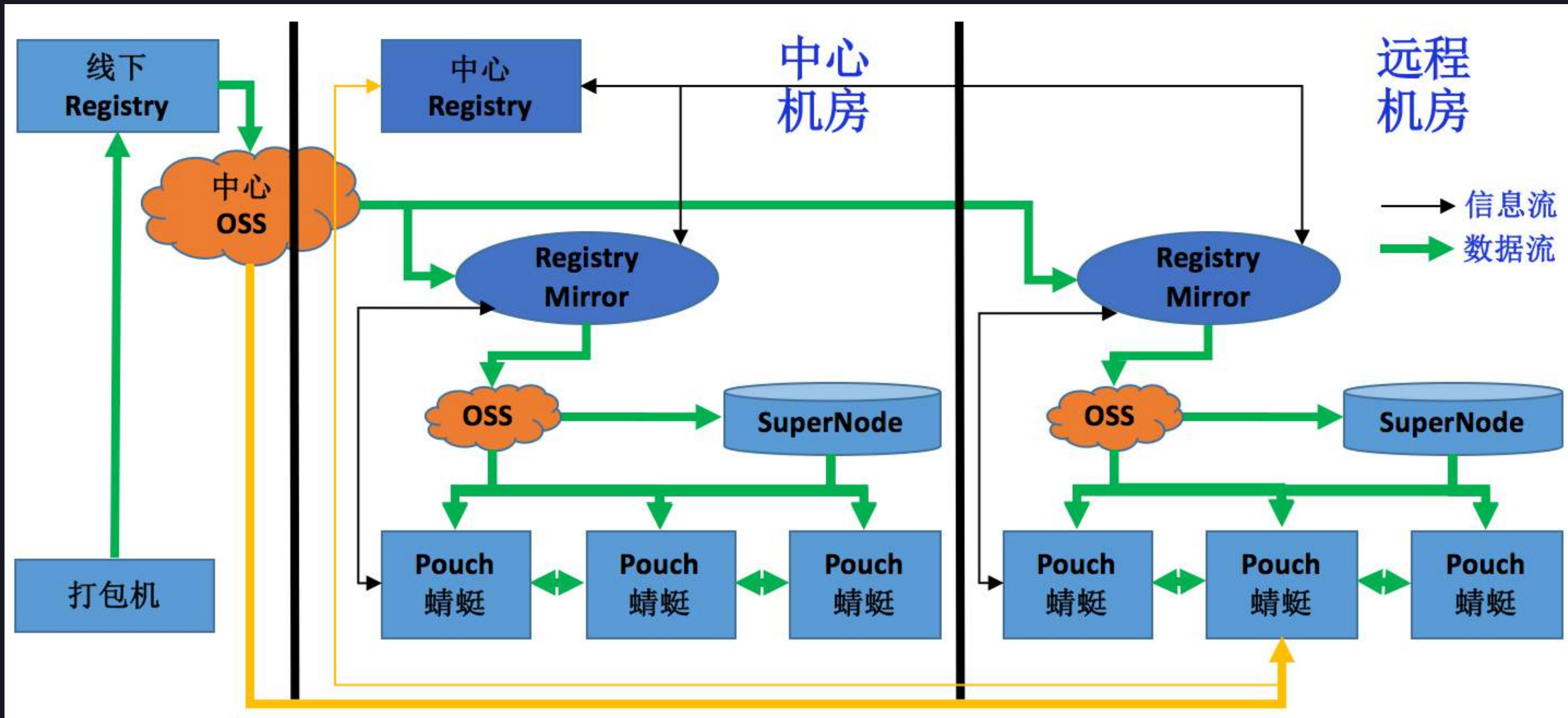
# Pouch 技术优势-更好的隔离性

- 传统容器的隔离维度：namesapce , cgroup
- 更优的容器可见性隔离：内核patch , lxcfs
- 额外隔离维度：磁盘，网络等：diskquota
- 基于Hypervisor的强容器隔离
  - runV
  - clear container

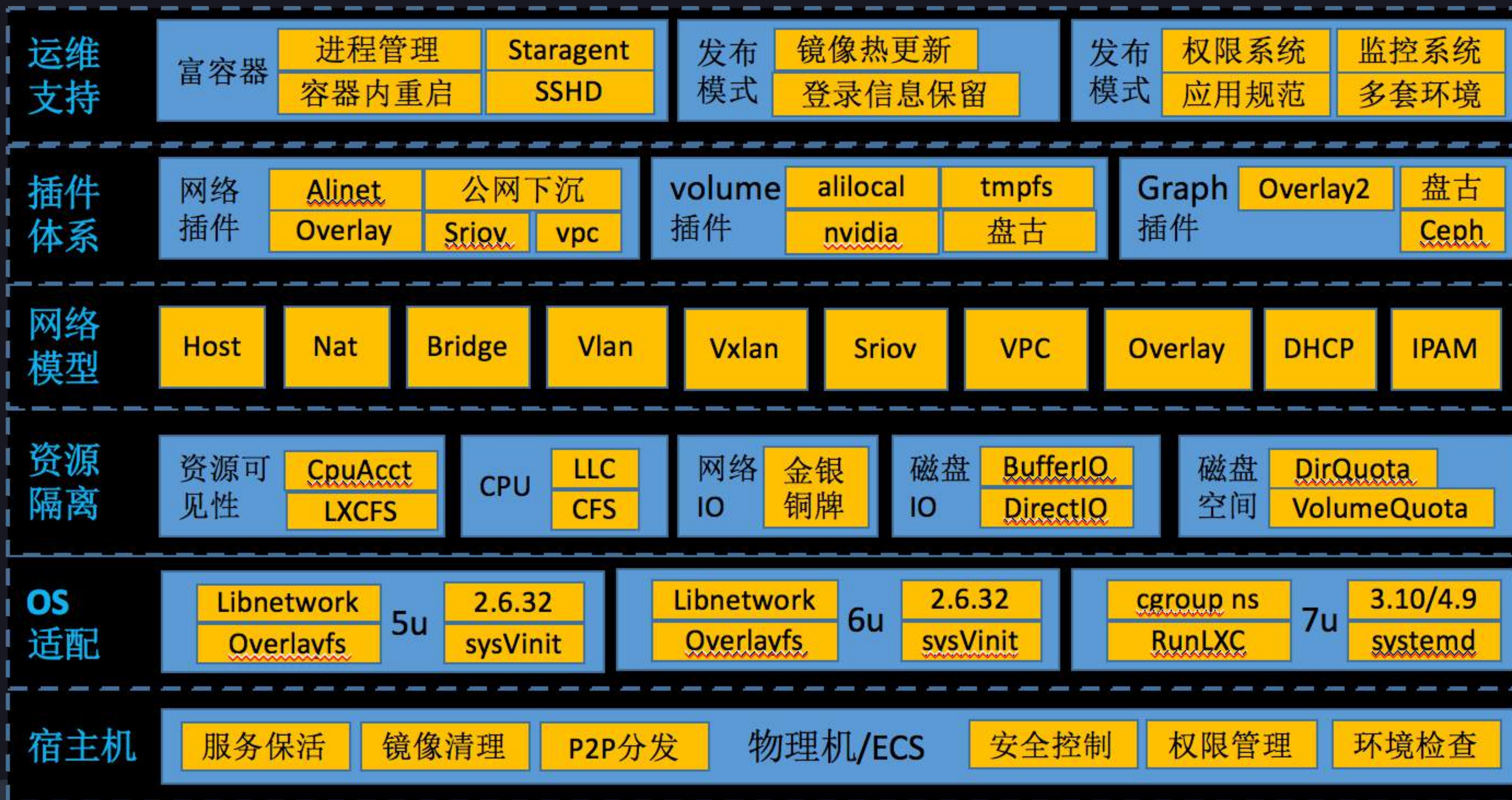
# Pouch 技术优势 - 隔离性支持混部



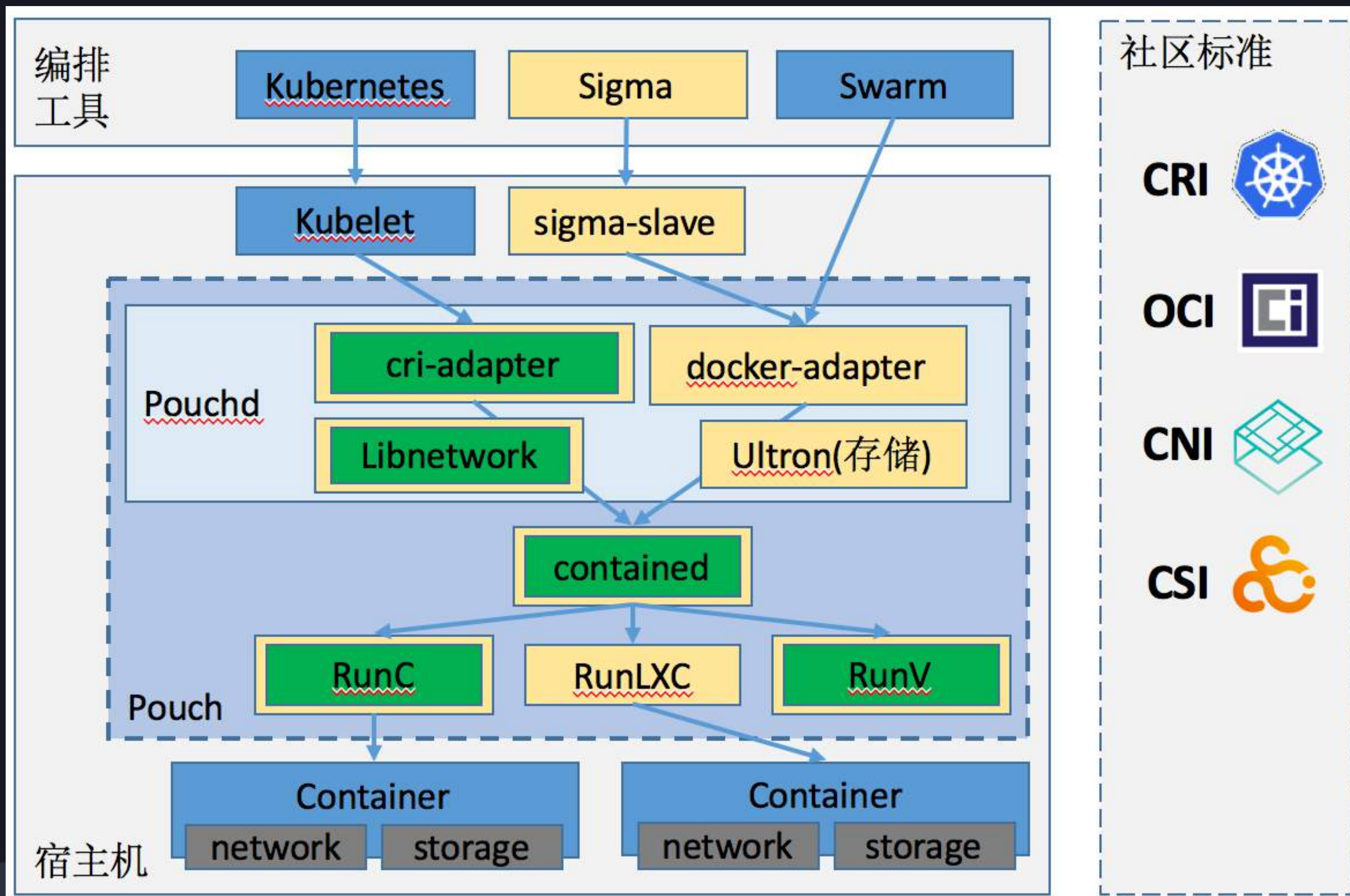
# Pouch 技术优势-镜像P2P分发



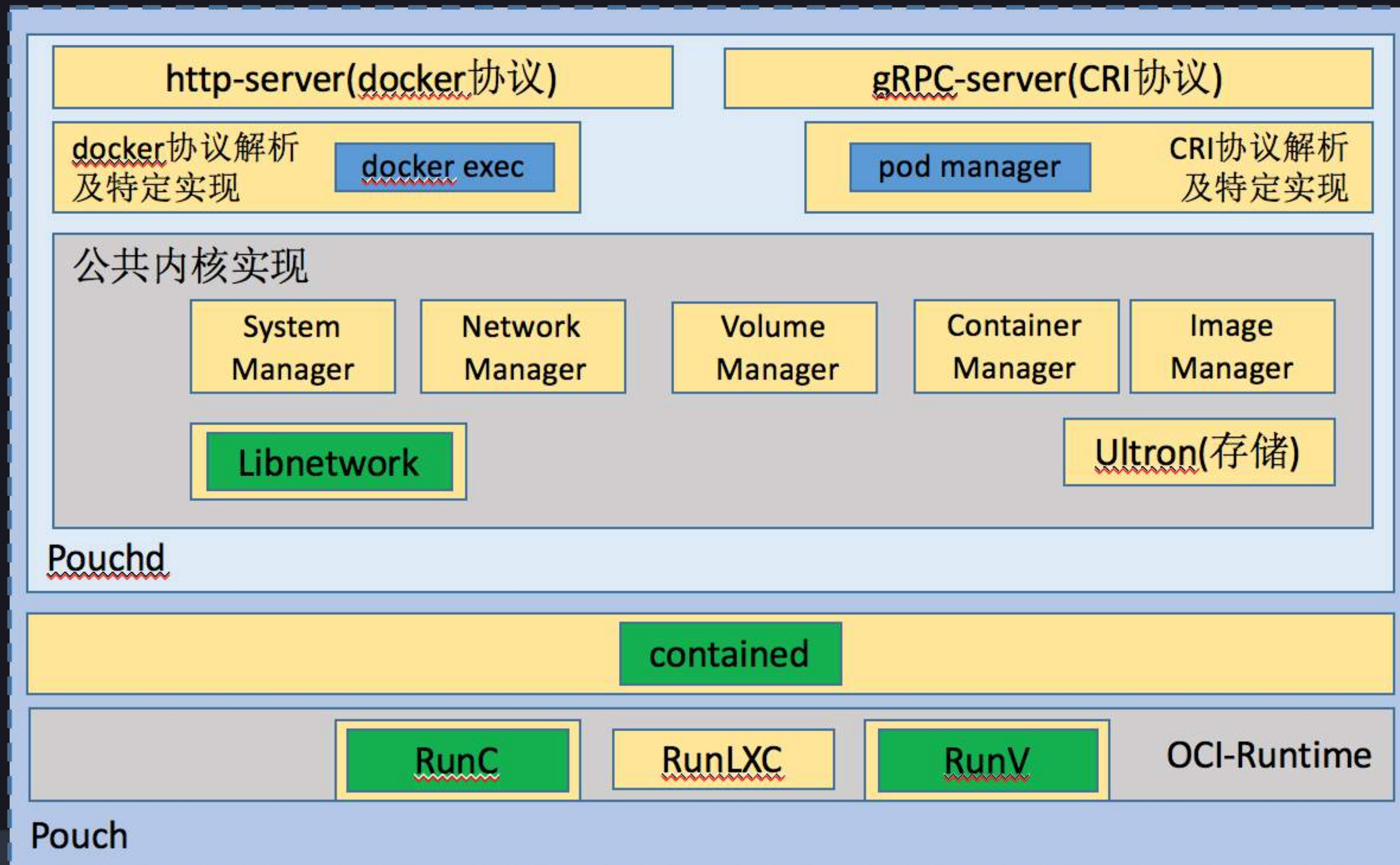
# Pouch 内部版本的体系结构



# Pouch 开源版本系统结构



# Pouch 开源版本系统结构



# Pouch 开源版本

欢迎参与：

<https://github.com/alibaba/pouch>

如何贡献：

<https://github.com/alibaba/pouch/blob/master/CONTRIBUTING.md>



# Pouch 开源计划



# THANK YOU

如有需求，欢迎至 [ 讲师交流会议室 ] 与我们的讲

  
**ArchSummit**  
全球架构师峰会 2017



阿里Sigma&Pouch林轩... 



扫一扫上面的二维码图案，加我微信