

大数据服务在云原生应用中的使用探索

北京亚信智慧数据科技有限公司

橘云产品线

叶鹏

当前大数据应用开发的困难点

如何更好的支撑大数据应用开发

基于ServiceBroker的大数据应用开发实践

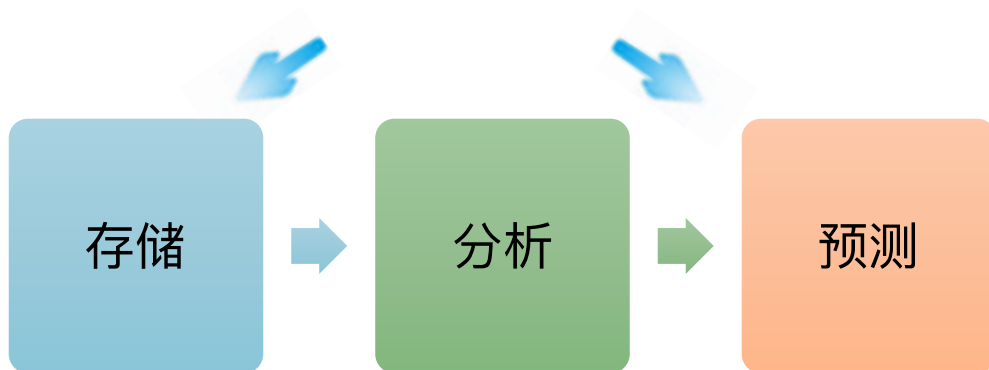
效果及存在问题

总结

少量共性

大数据应用

大量差异



GBASE

APACHE
STORM™

mongoDB

PostgreSQL

H₂O.ai



HBASE

MySQL

Spark

Apache
Solr

neo4j

redis

kafka



RabbitMQ



GREENPLUM
DATABASE



elasticsearch.

开发

给我的应用分用户了吗？

平台地址是多少？

能看到平台的控制台吗？

数据的权限还是不够呀！

运维

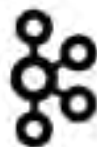
还没收到工单

昨天就邮件给你了

稍等，我再看看是啥问题

额，不能再大了

平台



kafka



mongoDB.



开发

学习特定平台

应用迁移困难

自主应用开困难

无法承载无状态
应用

运维

Report SaaS

ETL SaaS

ML SaaS

平台



kafka



mongoDB.



不完整版

有时间就能解决

开发流程
不畅
手工操作
多

应用托管
难度大
• 平台约
束

.....

Since 2000, 52% of the Fortune 500 are no longer on the list

The average lifespan of an S&P company dropped from 67 years in the 1920s to 15 years today

Transient Advantage

Fast? Safe? Scalable?



当前大数据应用开发的困难点

如何更好的支撑大数据应用开发

基于ServiceBroker的大数据应用开发实践

效果及存在问题

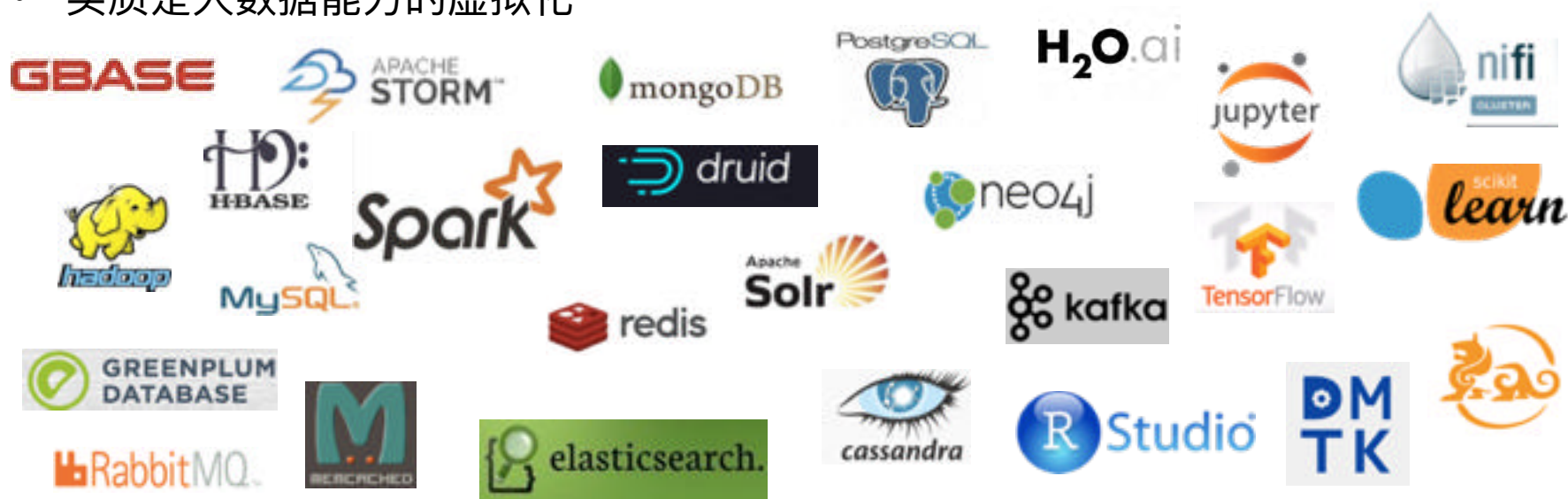
总结

“使应用具有较强适应性、敏捷性、弹性，并且可协作的一组持续改进的方法论和工具集”

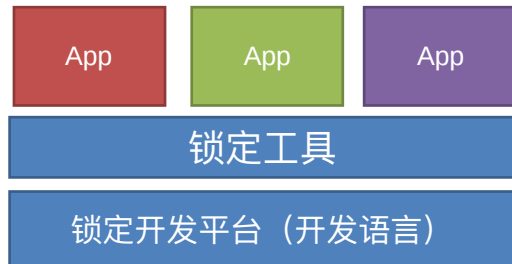
云原生工具集



- 必须采用标准化的集成方式
- 集成公有云能力和私有云已有的能力
- 实质是大数据能力的虚拟化



<http://docs.cloudfoundry.org/services/api.html>



- 仅限于使用平台方提供工具，如ETL、可视化
- 开发语言限制在平台提供的范围之内

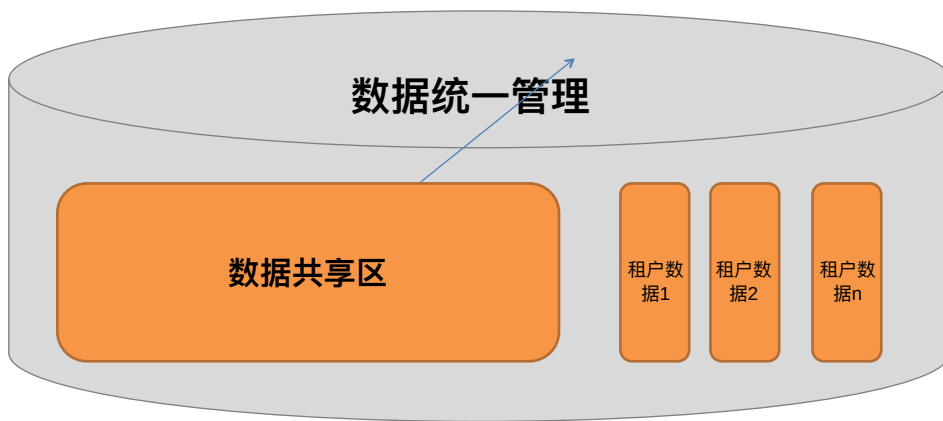
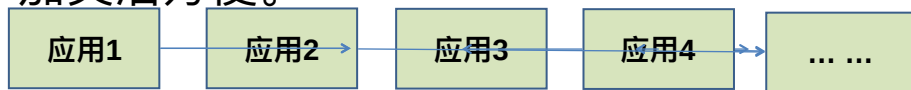


- 可利用已有或自有的任何工具，例如Jupyter
- 可以使用任何语言，如R
- 大数据的应用不仅限于报表和展示，个性化推荐、智能引导、嵌入式应用才是其大价值体现

数据也作为一种服务对外提供。

数据不多份存储，降低建设成本，避免碎片化。

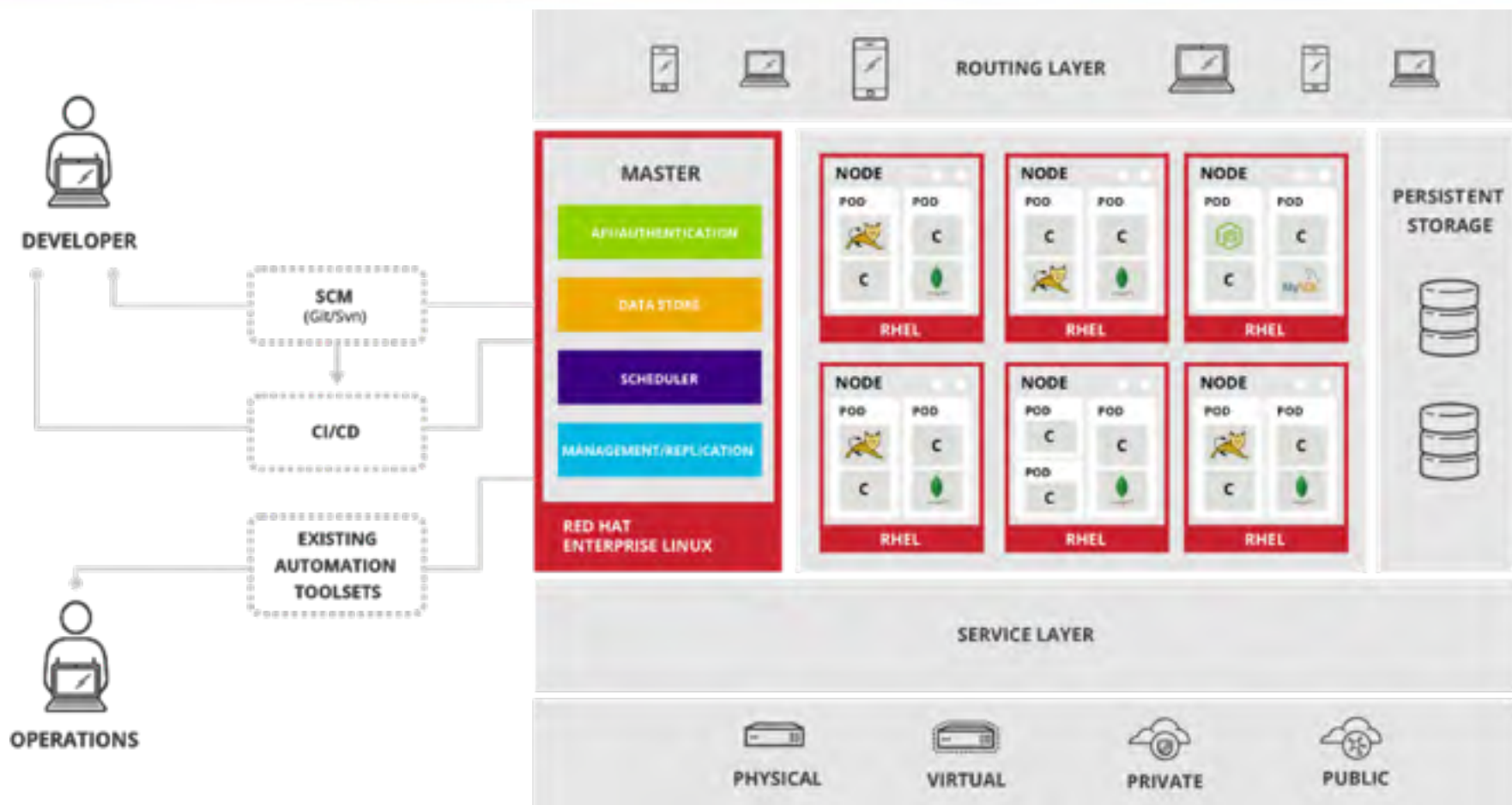
数据可以根据应用的需求划分多个租户区域，使得应用访问数据更加灵活方便。



- 数据标准
- 元数据
- 数据质量
- 数据治理



测试项目	平台部署	IAAS适配	
		部署能力	
		关键组件:	健康监控
			日志汇聚
			消息总线
			
	应用部署	GuestBook+mysql	
	内外部服务 框架	内部服务管理（编目、创建、绑定）	
		外部服务管理（编目、创建、绑定）	
		内部资源配额	资源配额类型/实现方式 资源隔离级别
		内外部服务发现	服务注册：域名、K-V、..... 动态扩容/负载均衡
		内外部认证鉴权	
		资源服务	RDB、NOSQL、文件系统、
		网络方案	容器间通信 跨主机容器通信
	应用框架	运行时/中间件	
		代码仓库	
		cicd	
		镜像打包	
		镜像发布	
		部署自动化能力	
		健康监控	
		日志汇聚	
	其他评估项	项目开发方式：协议、是否全部开源	
		组件开源方式：协议、是否全部开源	

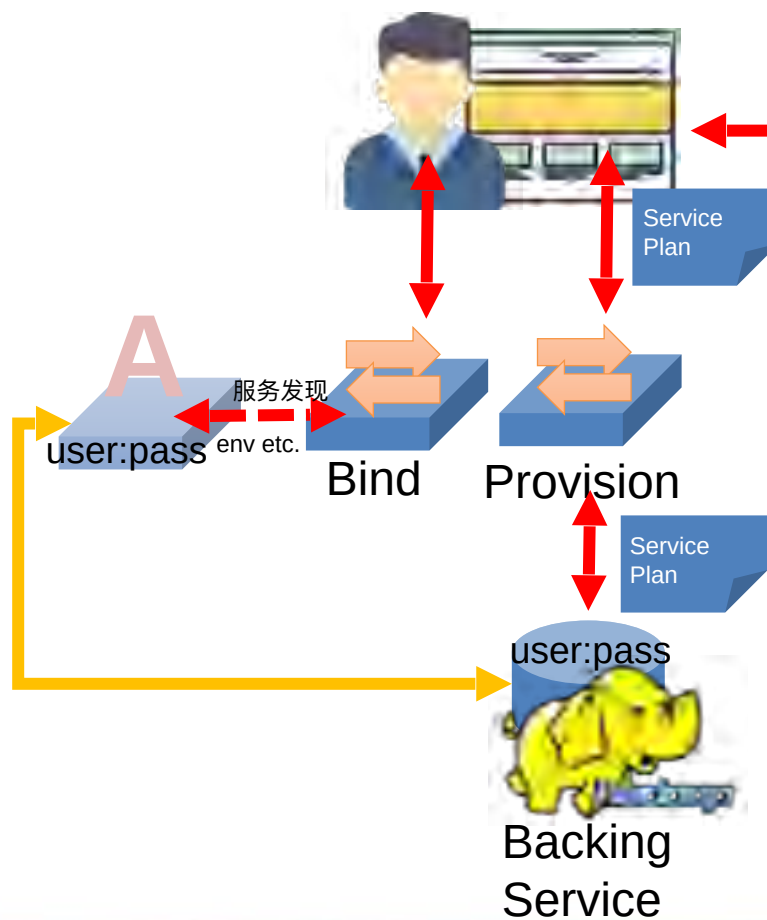


- CI/CD流程
- 内置镜像库
- Router
- ansible安装脚本
- 安全增强

<http://www.slideshare.net/SamuelTerburg/openshift-enterprise-31-vs-kubernetes>

- CloudFoundry ServiceBroker

- Cloudfoundry中后端服务与应用集成协议
- <http://docs.cloudfoundry.org/services/api.html>



- ✓ 业界事实标准，降低集成难度，复用开源实现，降低迁移难度。
- ✓ 以清晰的边界实现服务自治，实现了专业化地提供服务和便利地使用服务的分工。
- ✓ 以优雅的方式实现了大数据能力的虚拟化，提高资源使用率。
- ✓ 控制流和数据流分离，服务和服务可以打通，更加适合多样化大数据能力的集成。

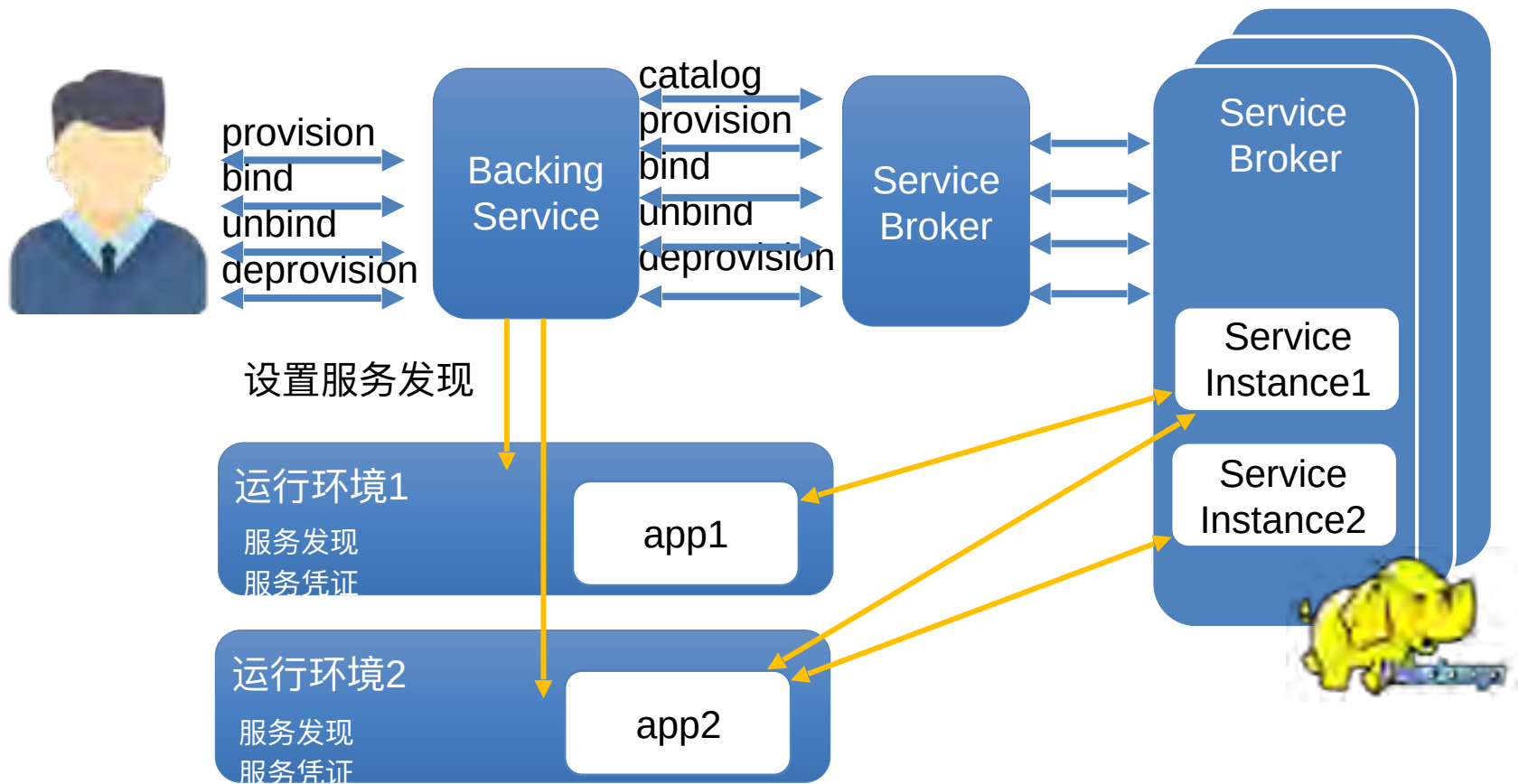
当前大数据应用开发的困难点

如何更好的支撑大数据应用开发

基于ServiceBroker的大数据应用开发实践

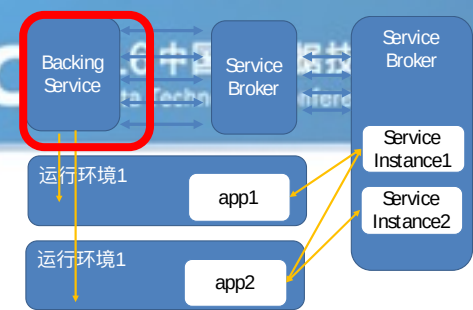
效果及存在问题

总结



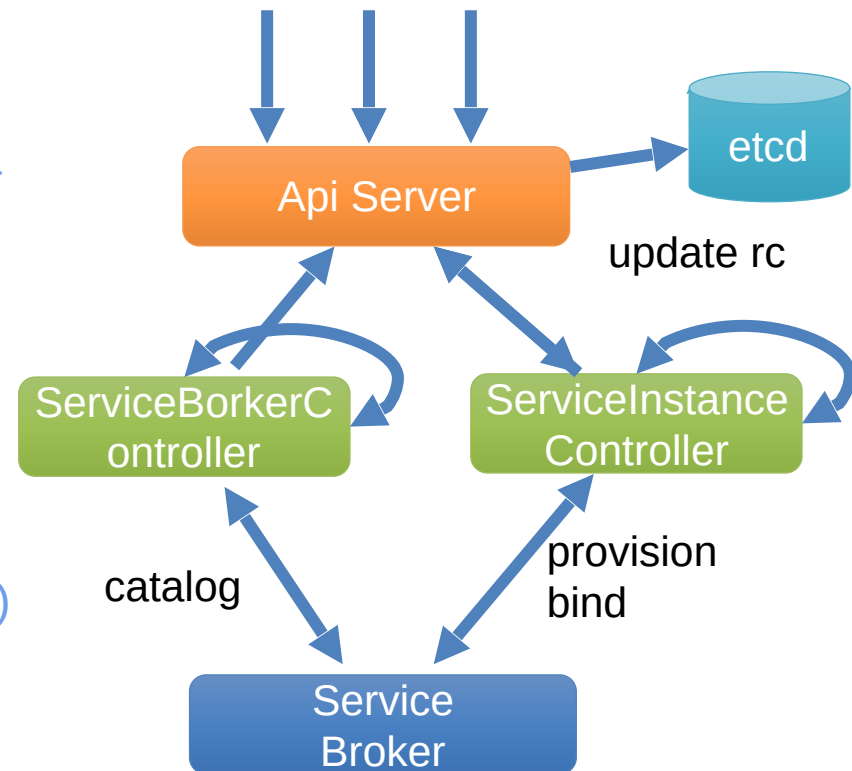
基于k8s的BackingService实现

BDTC



- type Service struct {...}
- type Pod struct {...}
- type ReplicationController struct {...}
- ...
- ...
- type ServiceBroker struct {...}
- type BackingService struct {...}
- type ServiceInstance struct {...}
- RunNodeController()
- RunScheduler()
- RunReplicationController()
- ...
- ...
- RunServiceBrokerController()
- RunServiceInstanceController()

create servicebroker
create serviceinstance

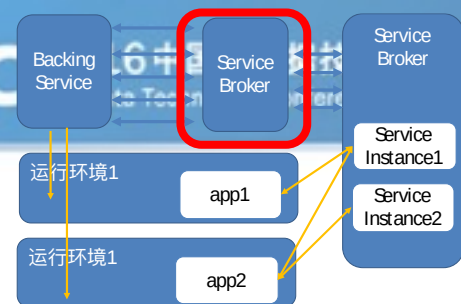


<https://github.com/asiainfoLDP/datafactory.git>

- ServiceBrokerController
 - Fetch catalog
(GET /v2/catalog)
- ServiceInstanceController
 - Provision instance
(PUT /v2/service_instances/:id)
 - Creating binding
(PUT /v2/service_instances/:id/service_bindings/:id)
 - Update RC
 - Remove binding
(DELETE /v2/service_instances/:id/service_bindings/:id)
 - Update RC
 - Remove instance
(DELETE /v2/service_instances/:id)

大数据服务ServiceBroker实现

BDTC



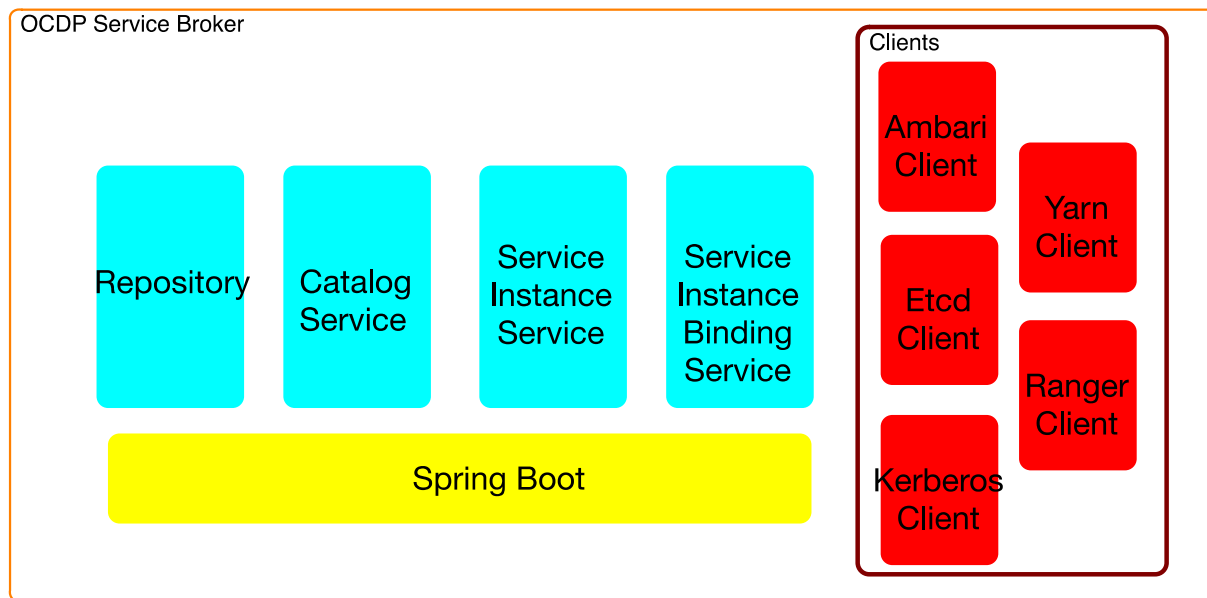
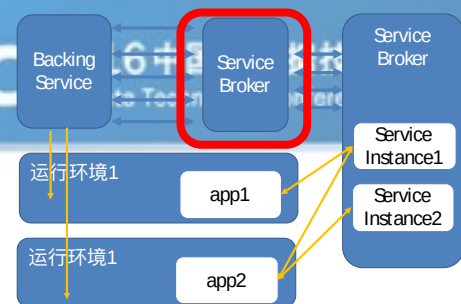
✓ 虚拟机



✓ 容器

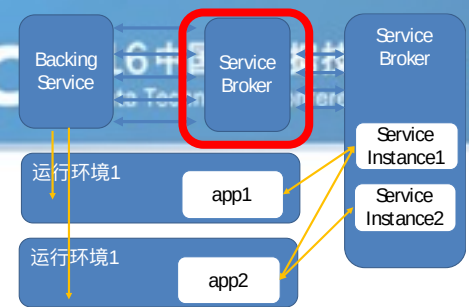


✓ 原生多租户

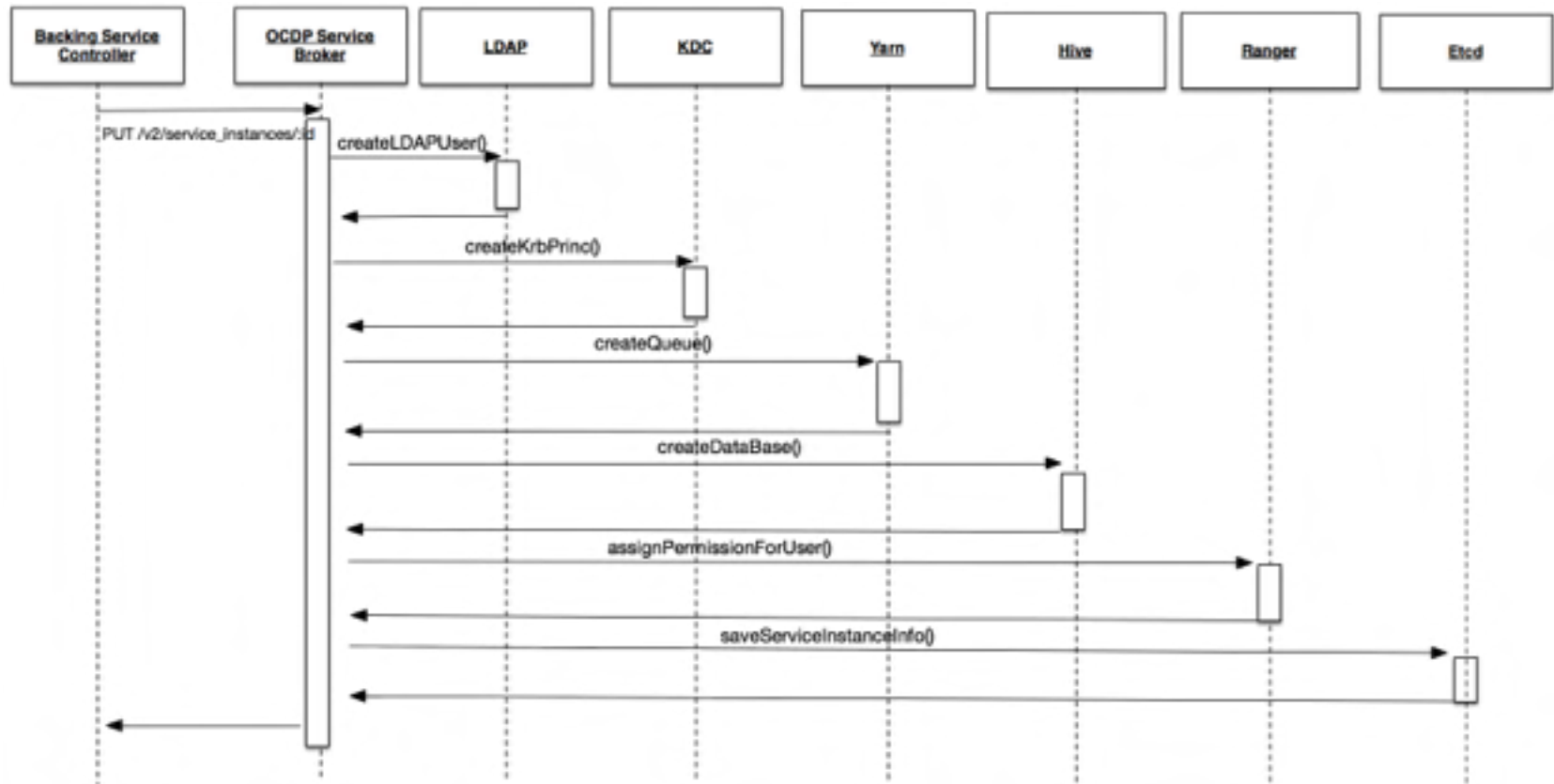


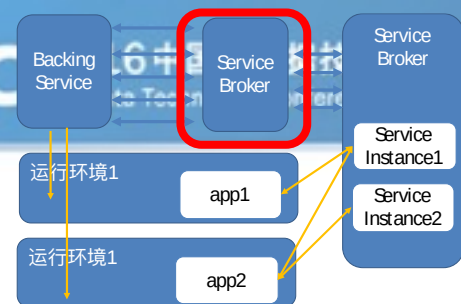
OSCP Service Broker可以划分为以下6：个部分

- Spring Boot: Spring Boot框架为service broker提供了实现REST服务
- Repository负责服务实例以及服务绑定实例的持久化中；
- Ranger鉴权
- Kerberos认证

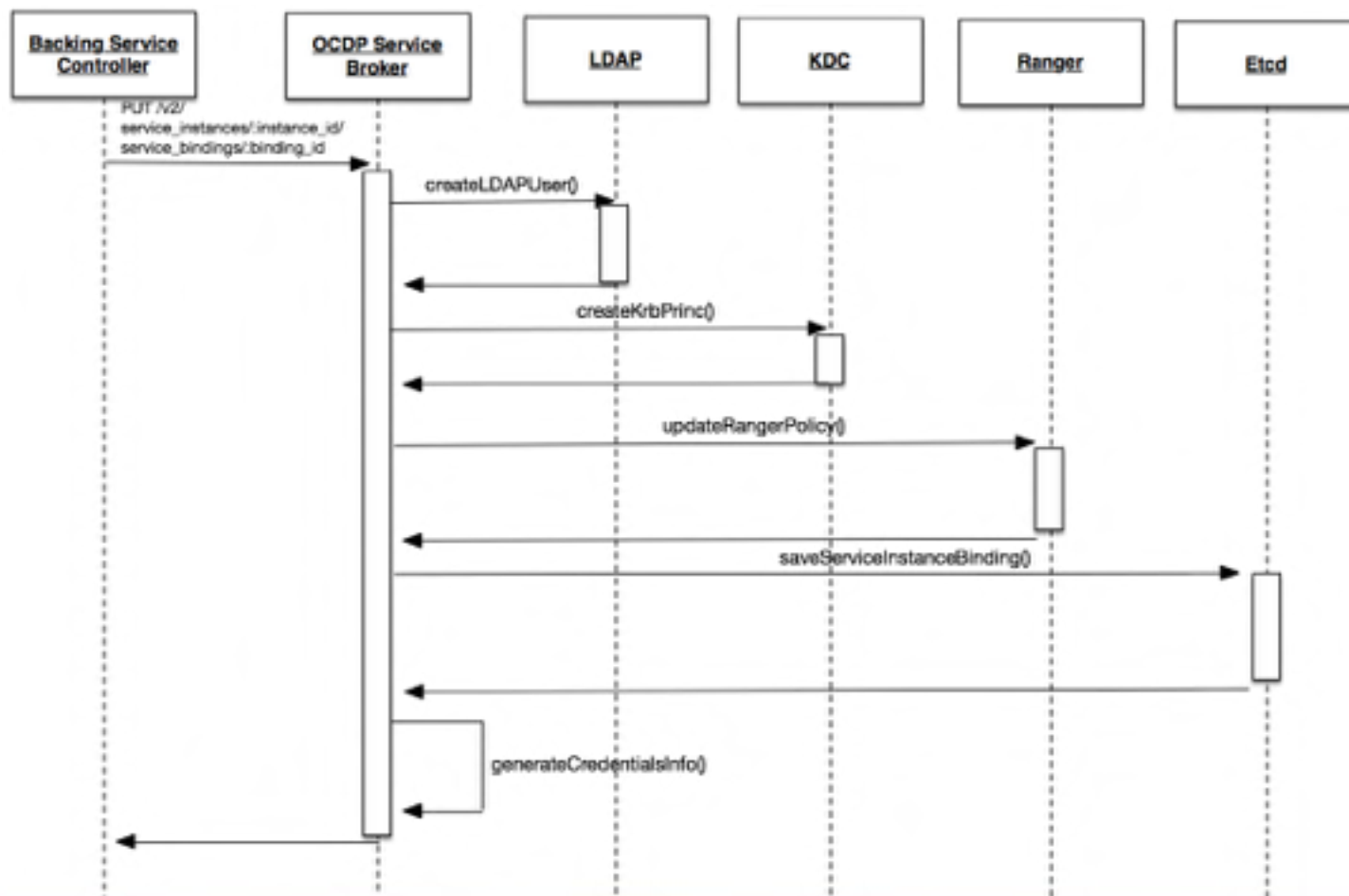


provision



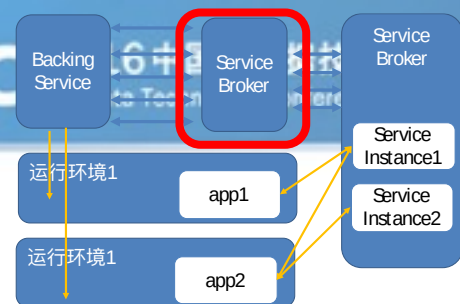


bind



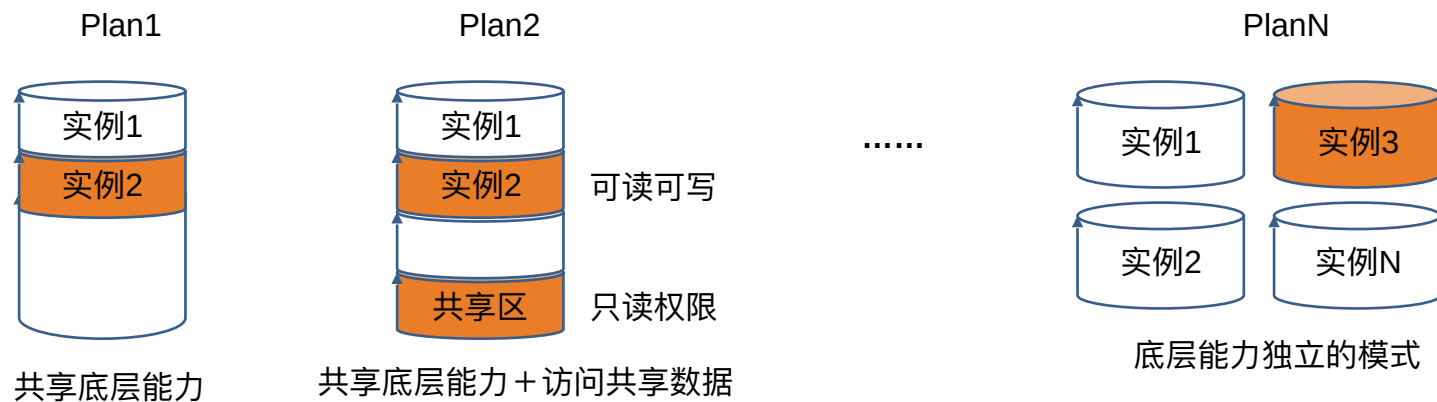
将数据本身作为服务进行管理

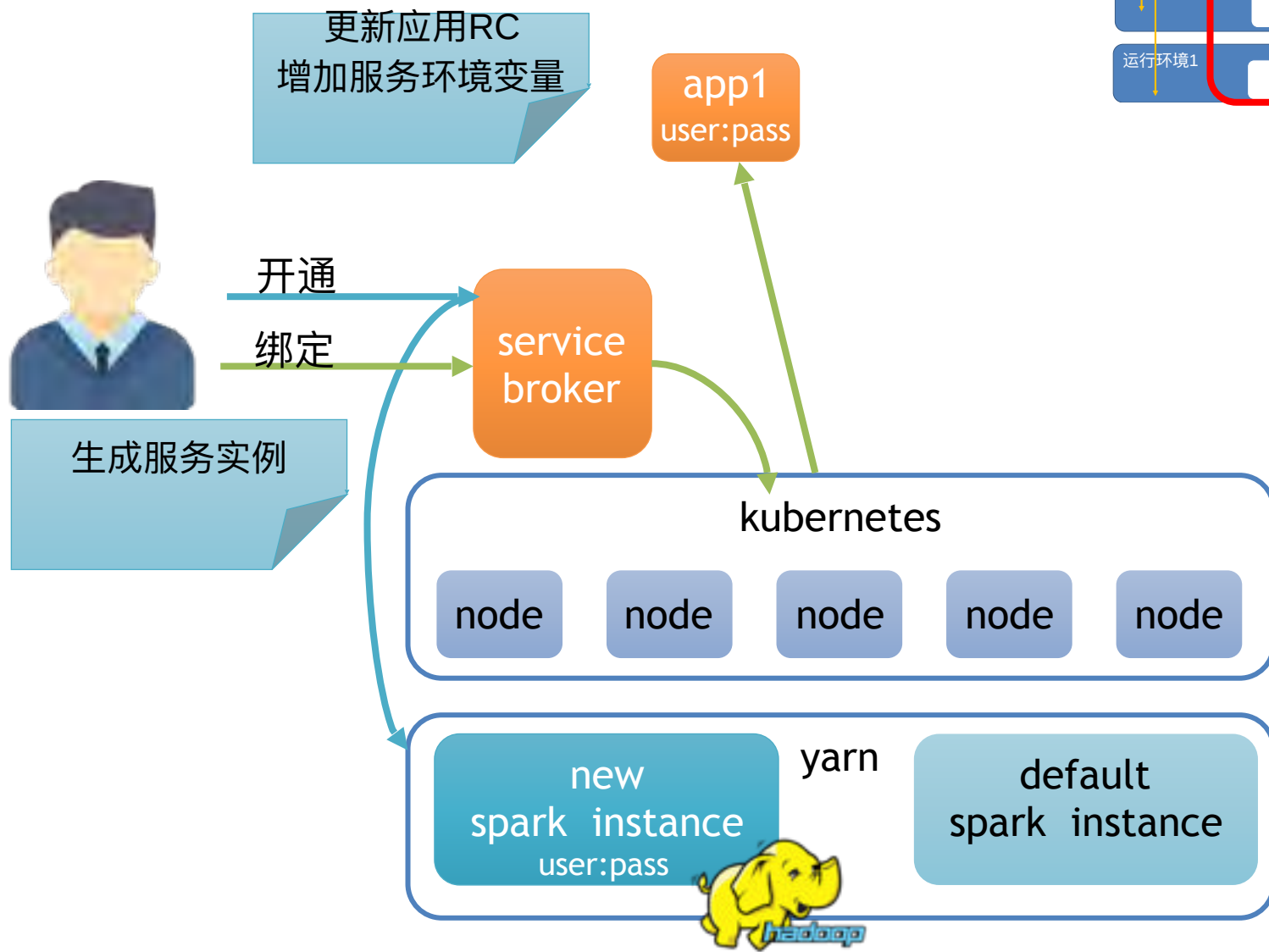
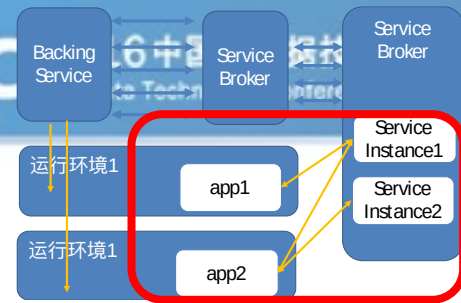
BDTC

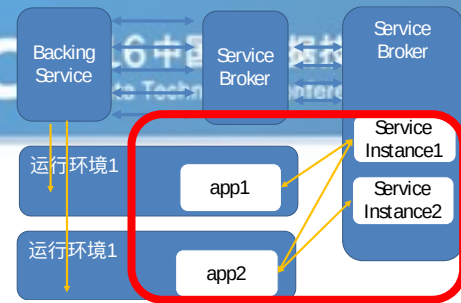
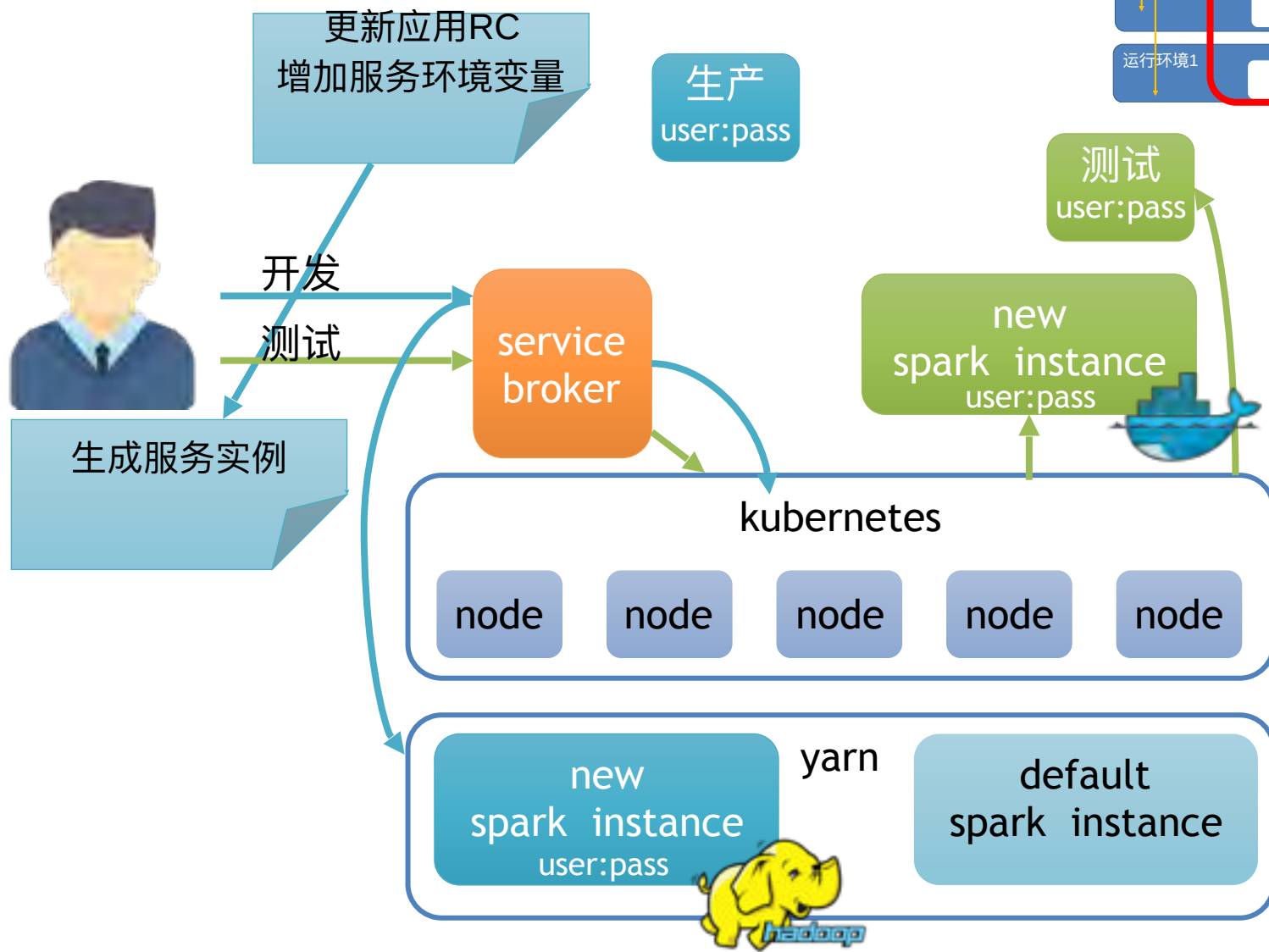


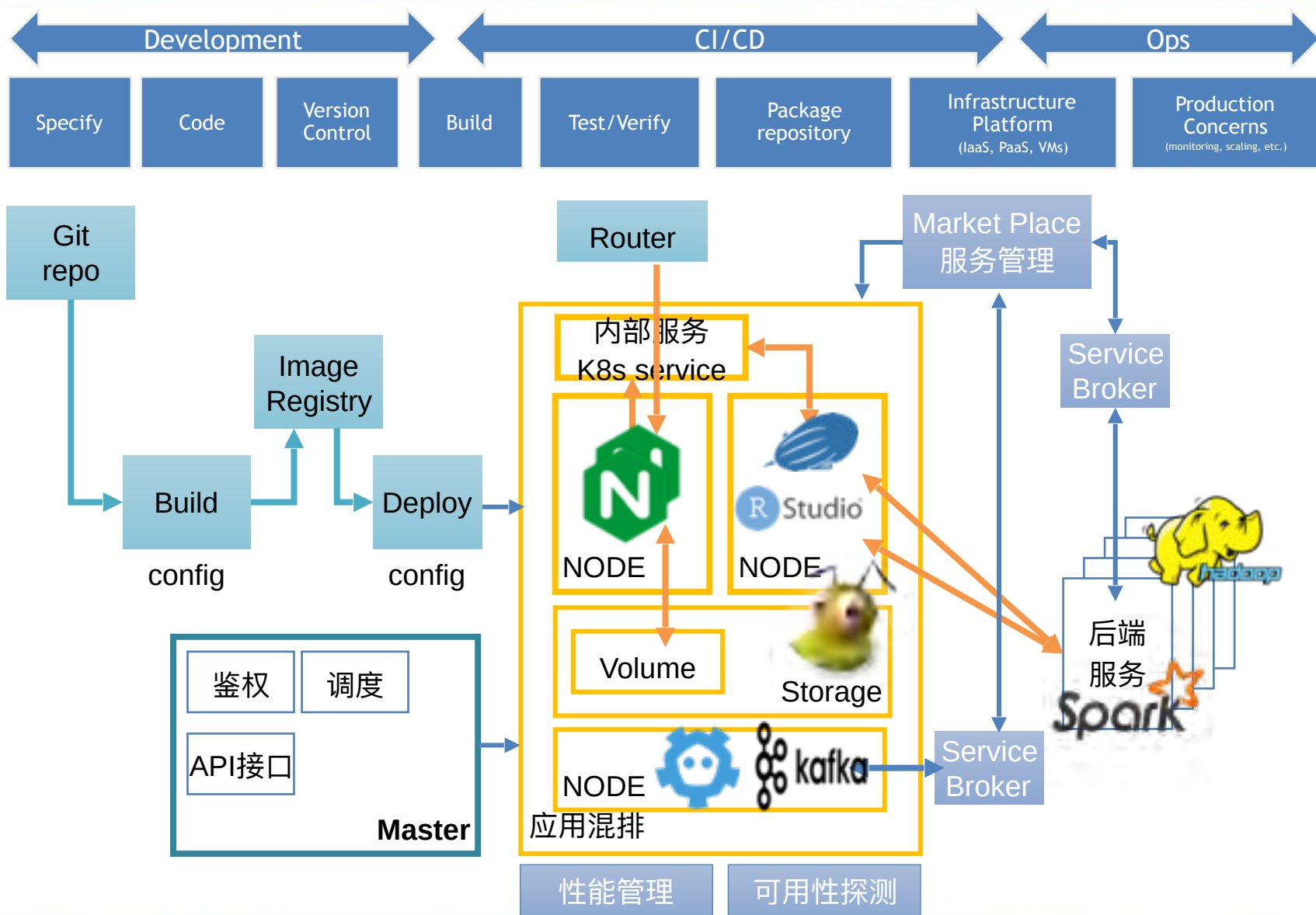
✓ 自助申请和管理

✓ 自动化生成实例









当前大数据应用开发的困难点

如何更好的支撑大数据应用开发

基于ServiceBroker的大数据应用开发实践

效果及存在问题

总结

 后端服务市场

 已申请的后端服务实例

 自定义后端服务实例

 数据集成服务实例

服务分类

数据库

计算

存储

消息

键值存储

服务总线

分布式协调

ETL

AI

其他

服务提供者

AI

亚信杨云

MapReduce

申请实例

Spark

申请实例

Asiainfo

Cassandra

申请实例

亚信杨云

HBase

申请实例

亚信杨云

HDFS

申请实例

亚信杨云

Hive

申请实例

 **公开数据集**

 数据集成服务

 DataHub平台介绍

数据分类

交通运输

天气信息

婚恋网站信息

居民消费价格数据

影视节目信息

房地产数据

手机APP信息

旅游信息

空气质量

股票信息

金融统计数据

数据标签

公开数据

演示数据

公开数据



空气质量

查看详情

演示数据



轨道交通

查看详情

公开数据



天气信息

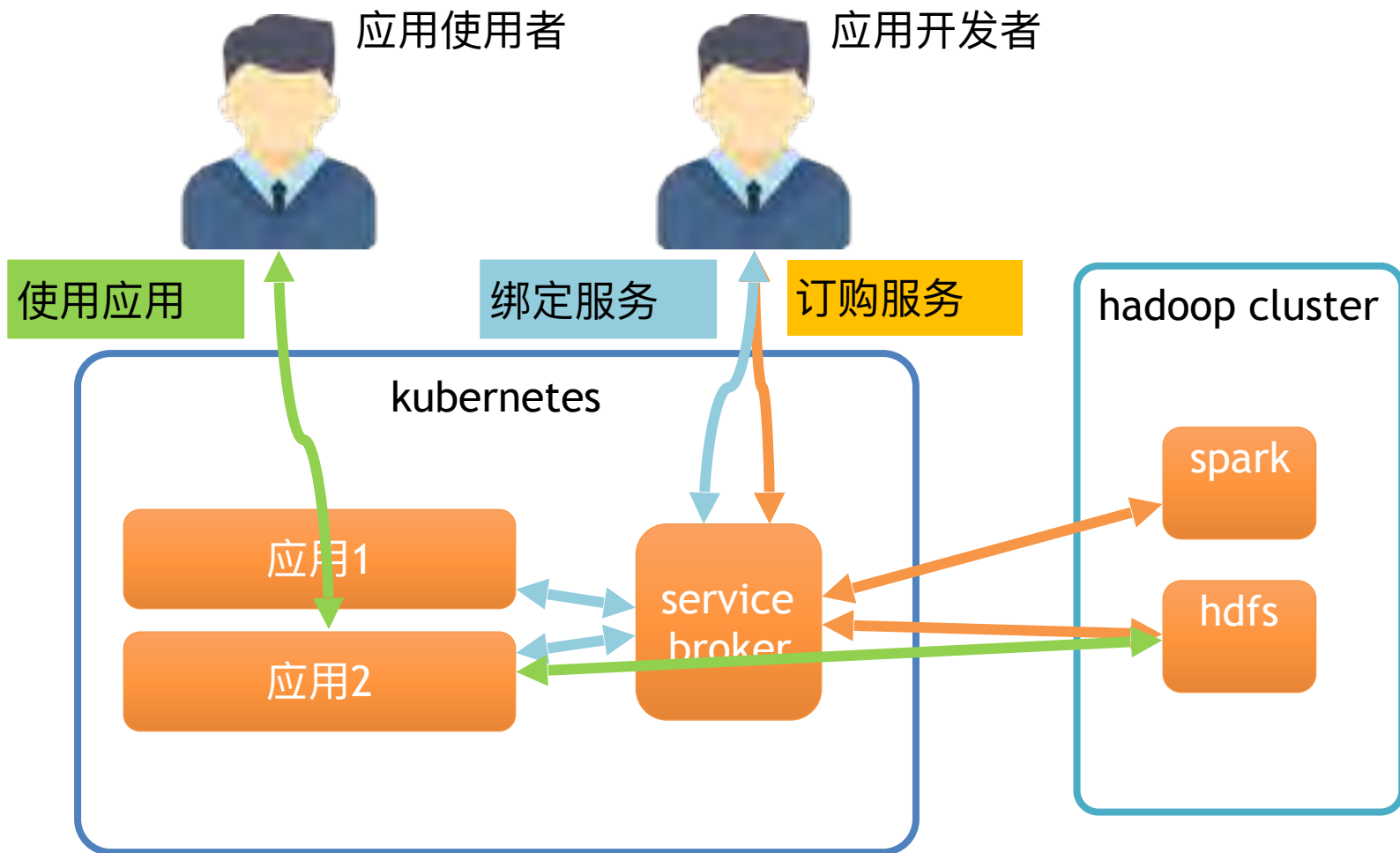
查看详情

公开数据



影视节目

查看详情



- 代码与kubernetes耦合度高
 - 使用kubernetes thirdparty resource对代码进行重构
 - 加入kubernetes service catalog SIG
- ServiceBroker模式对后端服务描述能力略弱
 - 通过Plan支持更灵活的资源申请
 - SaaS应用的集成

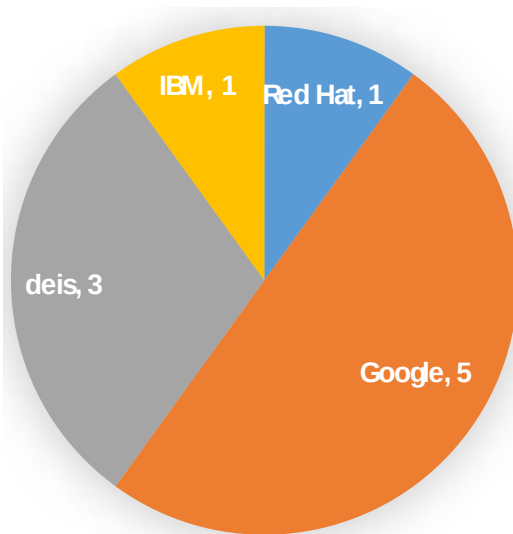
<https://github.com/kubernetes-incubator/service-catalog>

Purpose and Scope

The exact purpose of the SIG will grow over time, and as such, please see the various "version" folders under the [docs](#) directory for the exact list of use-cases, features and design decisions that have been agreed to.

However, this SIG is (initially) focused on an implementation of the Cloud Foundry Service Broker APIs. This would include the ability for Service Brokers to register themselves with a Kubernetes environment the same way they can with a Cloud Foundry environment. Additionally, we will be exploring what can be done to help Service developers when they want to expose their code via a Service Broker.

We are currently scoping our v1 milestone [here](#). Interested in contributing? Check out the [documentation](#)



Sep 11, 2016 - Dec 6, 2016

Contributions to master, excluding merge commits

Contributions: Commits ▼



当前大数据应用开发的困难点

如何更好的支撑大数据应用开发

基于ServiceBroker的大数据应用开发实践

效果及存在问题

总结



Most Build Software for Innovation and Differentiation

75%

By 2020, 75 percent of application purchases supporting digital business will be "build," not "buy."

Forecast Analysis: Enterprise Application Software, Worldwide, 2Q15 Update.

Gartner®

<https://serverless.com/framework/>

有一天不再被格子衫、双肩包束缚

BDTC 2016 中国大数据技术大会
Big Data Technology Conference 2016

We are Enabler for Innovation



谢谢

DataFoundry

lab.dataos.io

极速铸造云端应用 释放大数据之美



扫描二维码 关注更多精彩

BDTC 2016中国大数据技术大会
Big Data Technology Conference 2016