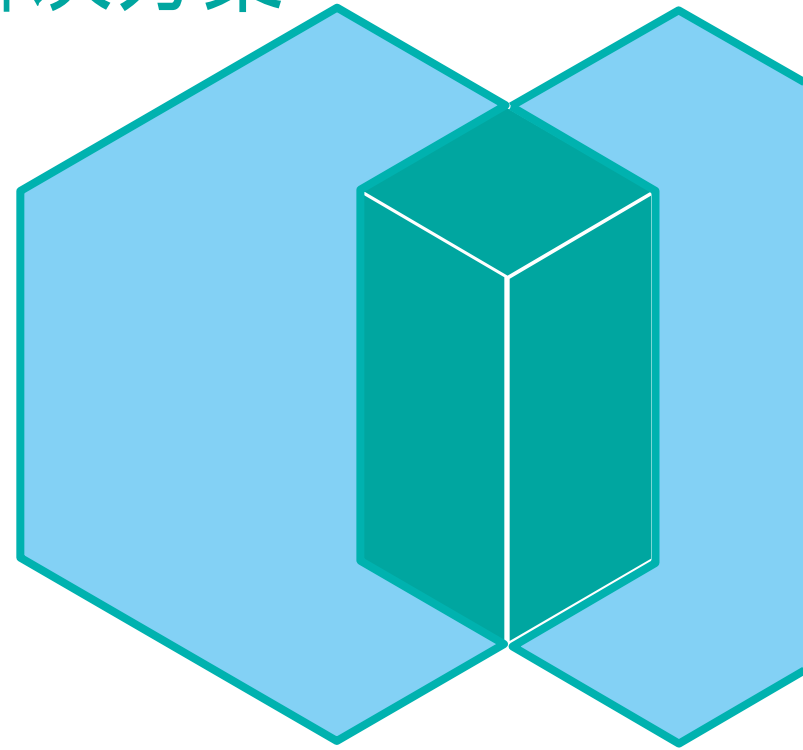
IBM DCOS 架构应用及解决方案

崔金, 解决方案架构师
cuijinc@cn.ibm.com



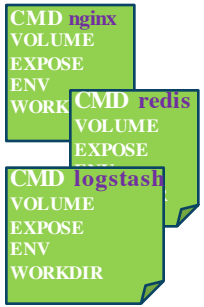
http://dss.cn.edst.ibm.com:81/campus/#/login?checkIn=Y&eventId=263
扫码注册&签到;若已经注册，用手机号再次登陆即完成签到。

目录

- 容器技术及Docker优点简介
- 数据中心操作系统DCOS介绍
- DCOS产品：Mesos、Kubernetes、IBM BlueDock
- DCOS应用场景及案例
- DCOS与云平台
- IBM在Docker、Mesos、Kubernetes社区的贡献
- BlueDock基本功能演示

Docker容器管理工具 – 功能

Build:
打包应用及依赖文件到容器，轻量
、操作系统平台无关，安全隔离



Docker build

Ship:
统一管理、分发高质量应用
应用分享

Docker Hub公共仓库有40多万个
“docker化”的应用

Docker run

Run:
快速部署、启动
一致的开发、测试、运维环境

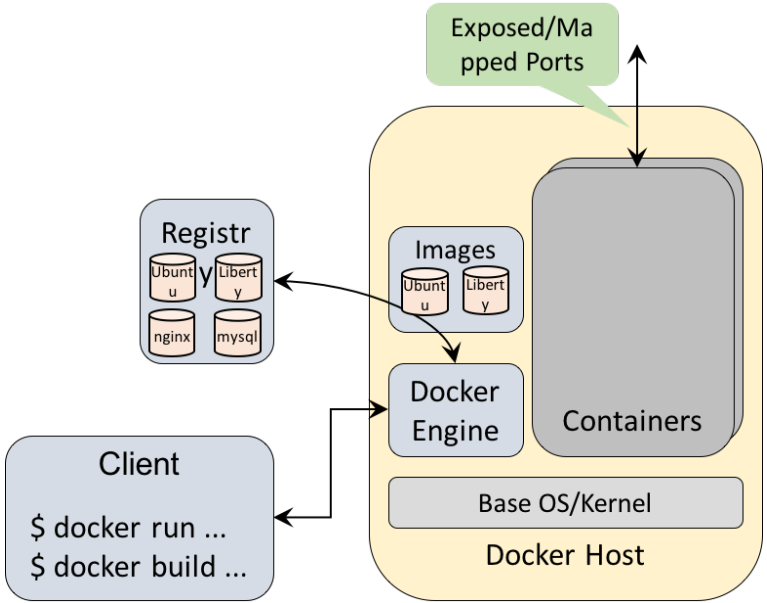
Docker Engine

基础架构：物理、虚拟机、存储、网络

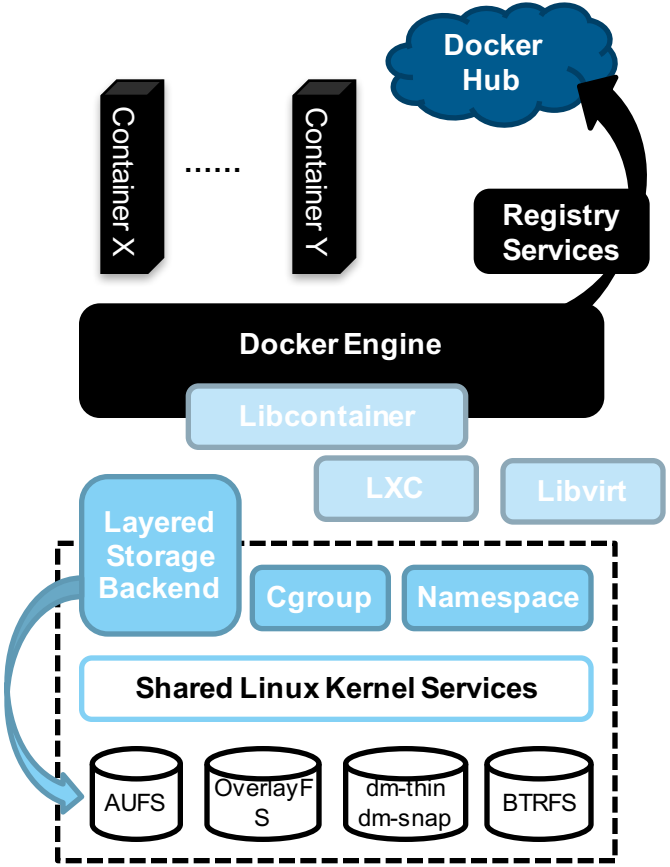
在Linux、MacOS、Windows上尝试安装使用Docker: https://docs.docker.com/engine/getstarted/

Docker 总体架构

- Docker Engine
  - 在主机上管理容器
  - 接收来自客户端的请求
    - REST API
  - 映射容器端口与主机端口
    - E.g. 80 → 3582
- 镜像
- Docker Client
  - 向Docker Engine发送命令
  - 创建并上传镜像
- Docker Registry
  - 镜像数据库



Docker container management tools – Runtime



Docker Hub (Cloud Registry Service)

- The docker registry as a cloud service, run by Docker, Inc.

Docker Engine

- CLI, utilities and daemon to run with.
- Docker Registry – a service to manage Docker images for storage and sharing.

Libcontainer / LXC (LinuX Container)

- Support to run a Linux system within another Linux system ("chroot() on steroids").
- Inside the box, it looks like a VM. Outside the box, it looks like normal processes.

Cgroup (Control Group)

- Limit, account and isolate resource usage (CPU, memory, dick I/O, etc.) of process groups.

Namespace

- Lightweight process virtualization (ipc, uts, pid, mount, network and user).

Layered Storage Backend

- Storage support for Docker's incremental image

Containers ≠ Docker

Containers are more than Docker

- "chroot jail"
  - Supported by most UNIX operating systems (1982?)
  - FreeBSD jail in 2000
- Solaris Containers & Solaris Zones
  - Solaris 10 in 2005
- AIX WPARs (Workload Partitions)
  - AIX 6.1(?) in 2007
- HPUX Containers (SRP)
  - Some time around 2007
- Linux LXC/LXD
  - Based on Google's cgroups/namespace work in 2006~2007
- Google containers in "borg"
  - Some time around 2006~2007?
- Docker、CoreOS Rocket/rkt、Mesos containerizer
- VM、JVM

Docker is more than Linux container

- The "libcontainer" extends Docker to other isolation mechanisms
  - For example, VMs with libvirt
  - That's how Docker can run on your Windows/Mac
- For some cases, it's actually an API/Standard
  - Native Docker on Windows
  - What's happening in Mesos/Kuberenetes

数据中心应用管理演进

应用开发：编码 测试

应用发布：集成 上线 运行

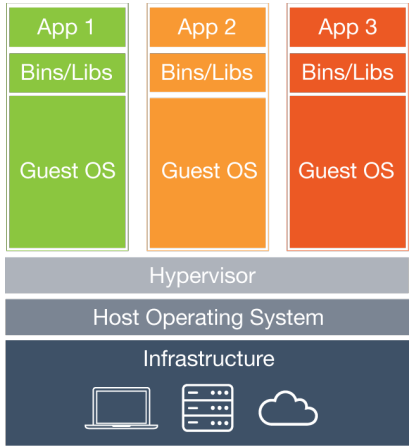
应用运维：容灾 扩容 迁移

应用目录：仓库 市场

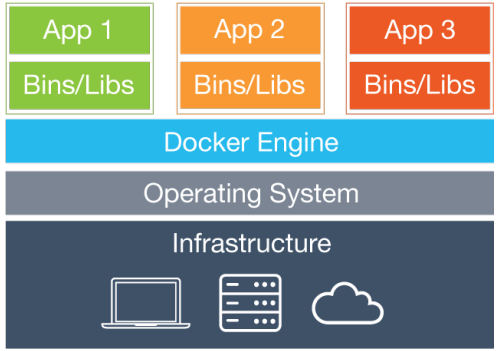
自动化 敏捷 弹性 高效

物理机
Bare Metal

基础架构云
IaaS



数据中心操作系统
DCOS



数据中心操作系统(DCOS)

- 资源管理能力
  - 相对于Linux、Windows等单机操作系统，DCOS能够管理分布在数据中心的异构的处理器、内存、网络、存储等资源
  - DCOS管理的容器可以运行在物理机或虚拟机，并同时弱化传统便于运维的基础架构云(IaaS)的价值，未来基于物理机的容器云将成为数据中心的演进方向
- 应用管理能力
  - 相对于当前多线程应用模型，未来应用将基于松耦合的微服务模型，运行在多容器中，通过RestAPI、MQ互联。 DCOS更够管理应用服务的生命周期、编排协同、分配调度等
  - DCOS可同时管理长运行服务、大数据分析及实时计算
- 背景及价值定位
  - 起源于Google服务使用的集群管理软件Borg，在Docker容器技术出来之后发展迅速，Mesos提升到DCOS概念，Platform EGO Grid OS是最早的企业级产品
  - 云计算三个阶段：服务器资源池化（虚拟化及IaaS），应用资源池化（PaaS），数据中心操作系统(DCOS)

数据中心操作系统(DCOS)

- DCOS特征
  - 分布式、微服务、容器化、无状态、云原生（Cloud Native）、动态伸缩、负载均衡、高可用(self-healing)、服务自动发现、滚动升级和回滚、灰度发布、持续运营（Devops, CI/CD）
- 应用场景
  - 当前主要应用于移动、互联网应用，如App抢红包、秒杀等，未来应用场景广泛，如CRM, BOSS，Core business应用
  - 当前发展瓶颈在DCOS技术的成熟度及软件微服务架构改造的难度
- 主流提供者
  - Google Kubernetes: Google数据中心集群管理Borg/Omega的开源简化版，并提供Google Cloud Platform公有云服务；擅长编排
  - Apache Mesos: 2009年加州大学伯克利分校开发的开源集群管理软件，随后成立Mesosphere初创公司提供商业DCOS软件及服务；架构清晰、擅长资源管理
  - Docker Swarm: Docker主导的容器资源调度和业务编排开源项目；简单易用
  - IBM Conductor with Container （CwC）: 基于Mesos、Kubernetes、Docker、 EGO、ASC具有丰富的资源管理、调度编排、并稳定高效

完整的数据中心操作系统需要的功能

- 自动装箱负载

根据容器的资源需求和约束条件自动放置到合适的主机上。混合调度关键负载和尽量运行负载以提高利用率，节省资源

- 自容错

重启失效的容器，替换并重新调度失效节点上的容器，根据用户自定义健康检查程序杀死没有响应的容器

- 水平扩展

使用简单的命令或图形界面触发扩容或收缩应用服务。或者根据CPU利用率自动伸缩。

- 服务发现及负载均衡

集成网络及DNS管理。为容器集合分配自己的IP及DNS名字，并可以在容器之间做负载均衡

- 自动发布和回滚

在保证正常访问的同时渐进的发布新版本应用。新应用有问题时可以回滚到前面的版本。

- 存储编排

自动mount容器卷，不管是本地文件系统还是GCP、AWS等网盘存储，或者NAS，如NFS、iSCSI、Gluster、Ceph、Cinder、Flocker等

- 安全及配置管理

部署和更新安全及应用配置无需重新创建镜像或在配置中暴露安全信息，如token，key等信息

- 批处理任务执行

除管理服务类型应用外，还可以管理批处理及持续集成（CI）应用

- 多租户

多层级的树形租户及资源分配结构

- 高级调度策略

优先级、资源共享抢占、借还、基于时间的调度，智能回收

- 应用集成、平台集成

资源、负载、镜像管理平台集成
应用容器化

- 用户管理

LDAP，Kerberos，Keystone集成

- 扩展性

扩展到数万节点

- 可用性

管理节点、计算节点无单点故障

- 异构

支持异构OS及CPU类型

- 可管理性

资源、容器监控、告警及图形展示

完整的数据中心操作系统需要的功能

Container Orchestration - Kubernetes

- 自动装箱负载

根据容器的资源需求和约束条件自动放置到合适的主机上。混合调度关键负载和尽量运行负载以提高利用率，节省资源

- 自容错

重启失效的容器，替换并重新调度失效节点上的容器，根据用户自定义健康检查程序杀死没有响应的容器

- 水平扩展

使用简单的命令或图形界面触发扩容或收缩应用服务。或者根据CPU利用率自动伸缩。

- 服务发现及负载均衡

集成网络及DNS管理。为容器集合分配自己的IP及DNS名字，并可以在容器之间做负载均衡

- 自动发布和回滚

在保证正常访问的同时渐进的发布新版本应用。新应用有问题时可以回滚到前面的版本。

- 存储编排

自动mount容器卷，不管是本地文件系统还是GCP、AWS等网盘存储，或者NAS，如NFS、iSCSI、Gluster、Ceph、Cinder、Flocker等

- 安全及配置管理

部署和更新安全及应用配置无需重新创建镜像或在配置中暴露安全信息，如token，key等信息

- 批处理任务执行

除管理服务类型应用外，还可以管理批处理及持续集成（CI）应用

Integration & Management

- 多租户

多层级的树形租户及资源分配结构

- 高级调度策略

[illegible]资源、[illegible]还[illegible]于时间的调度，智能回收

- 应用集成、平台集成

[illegible]应用容器化

- 用户管理

LDAP，Kerbose，Kestone集成

Resource Pooling - Mesos

- 扩展性

扩展到数万节点

- 可用性

[illegible]节点、计算节点无单点故障

- 异构

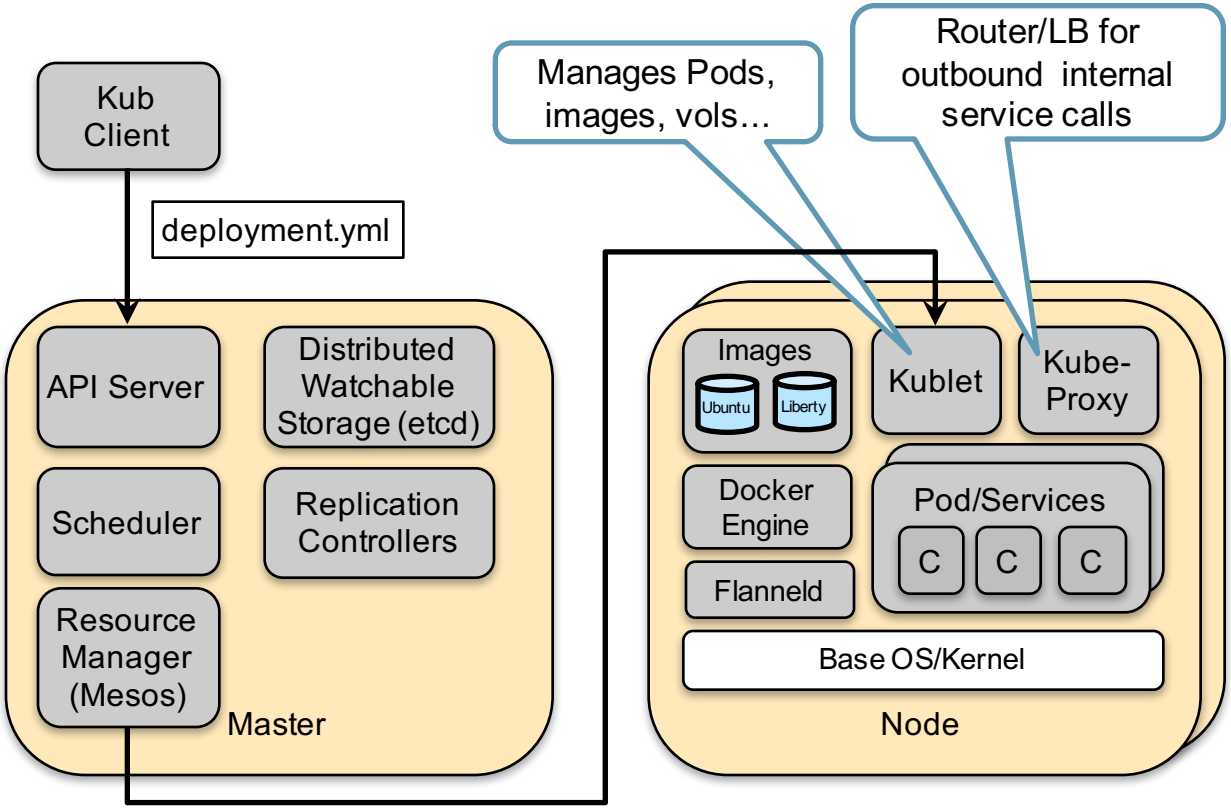
支持异构OS、CPU类型

- 可管理性

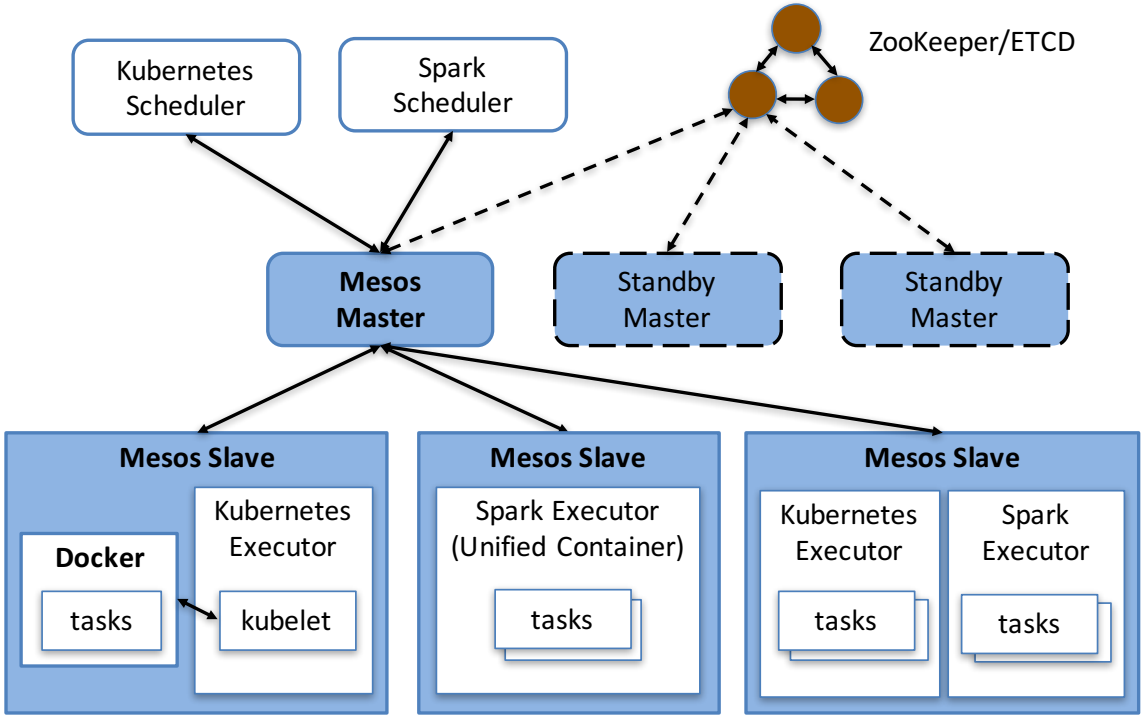
资源、容器监控、告警及图形展示

Kubernetes

- Kube-Proxy discovers services, load balance
- Pod is a group of one or more containers which share storage and IP
- RC makes sure Pods always up and available, rolling update
- Flanneld provides SDN

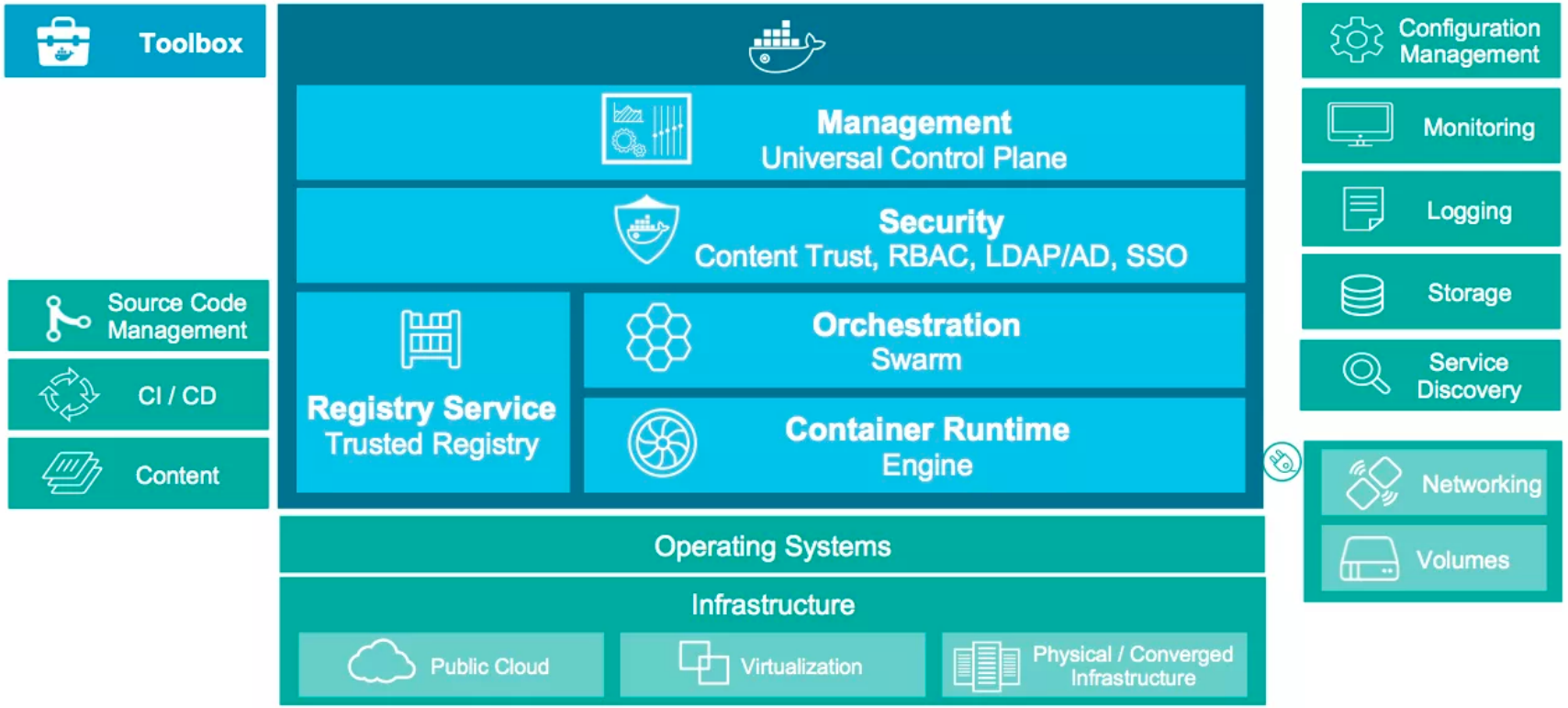


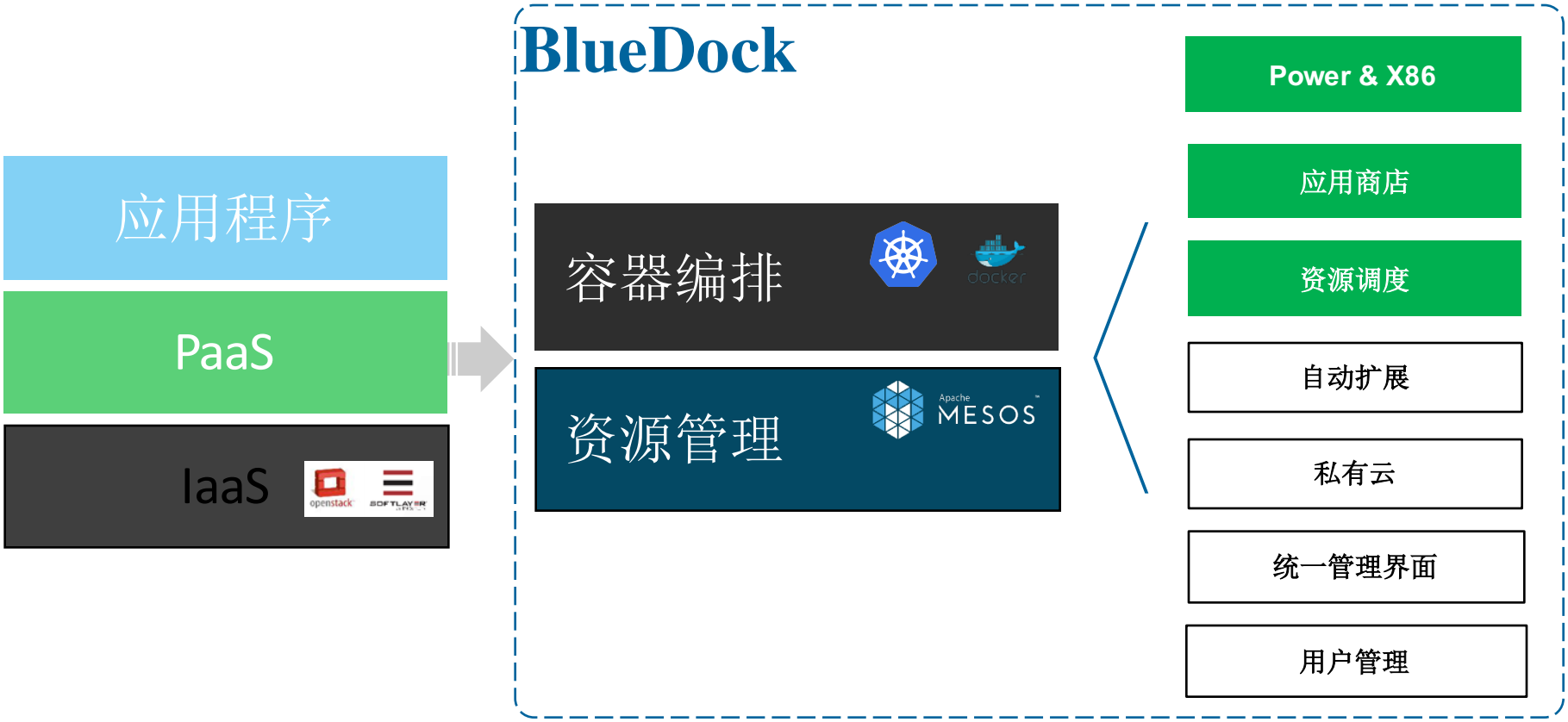
Mesos



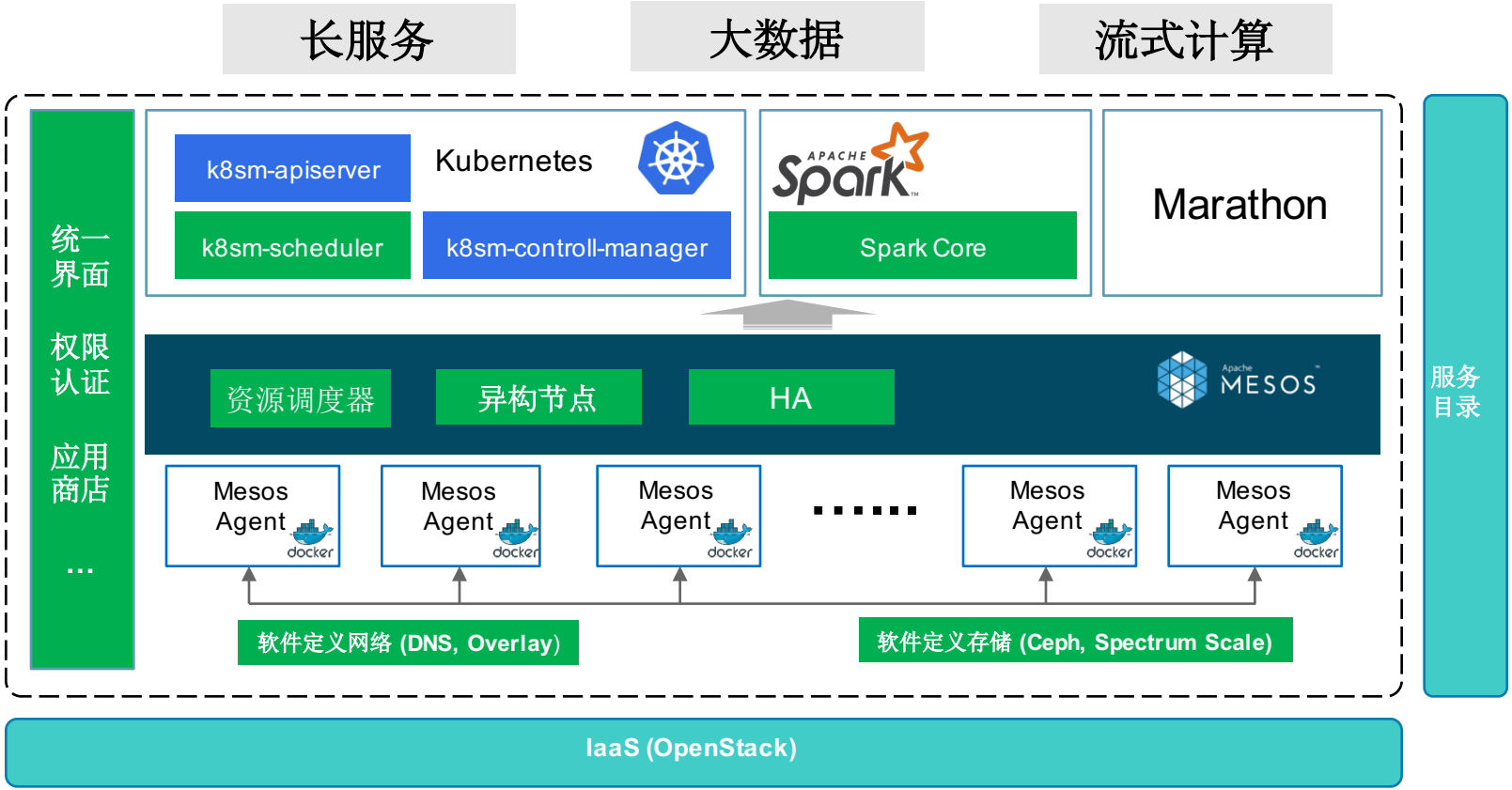
- **Apache** 开源项目
- 关注于资源管理
- 支持多种框架
- 两层调度
- **Twitter/Apple** 有上万节点的集群

Docker DataCenter



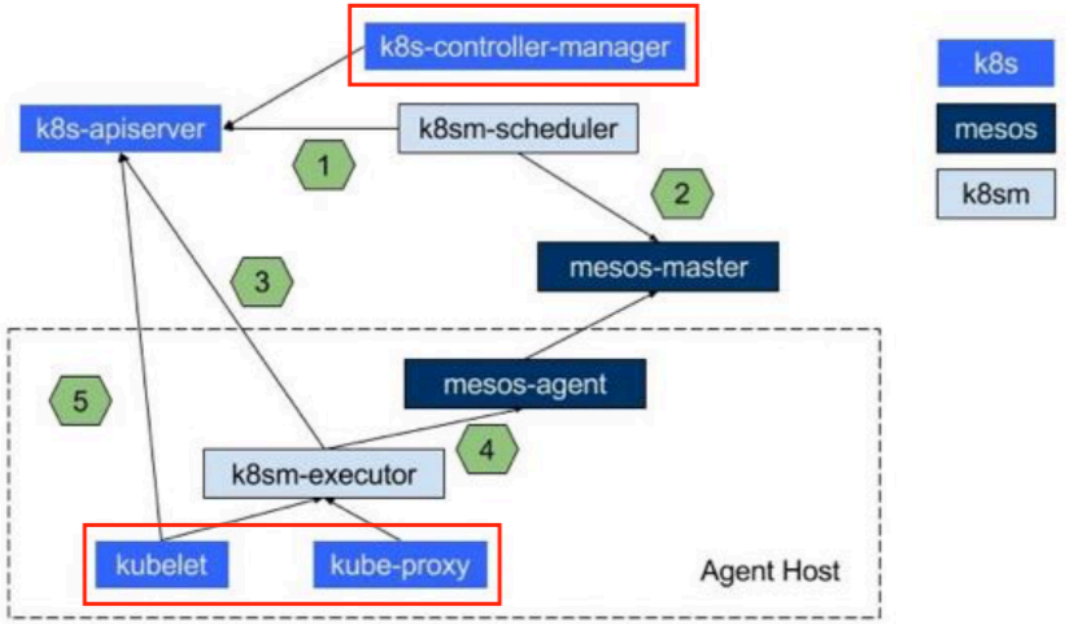

应用程序
PaaS
IaaS
openstack
SOFTLAYER
BlueDock
容器编排
docker
资源管理
Apache
MESOS
Power & X86
应用商店
资源调度
自动扩展
私有云
统一管理界面
用户管理

企业级容器平台 - Conductor with Container（绿色部分为IBM增强）

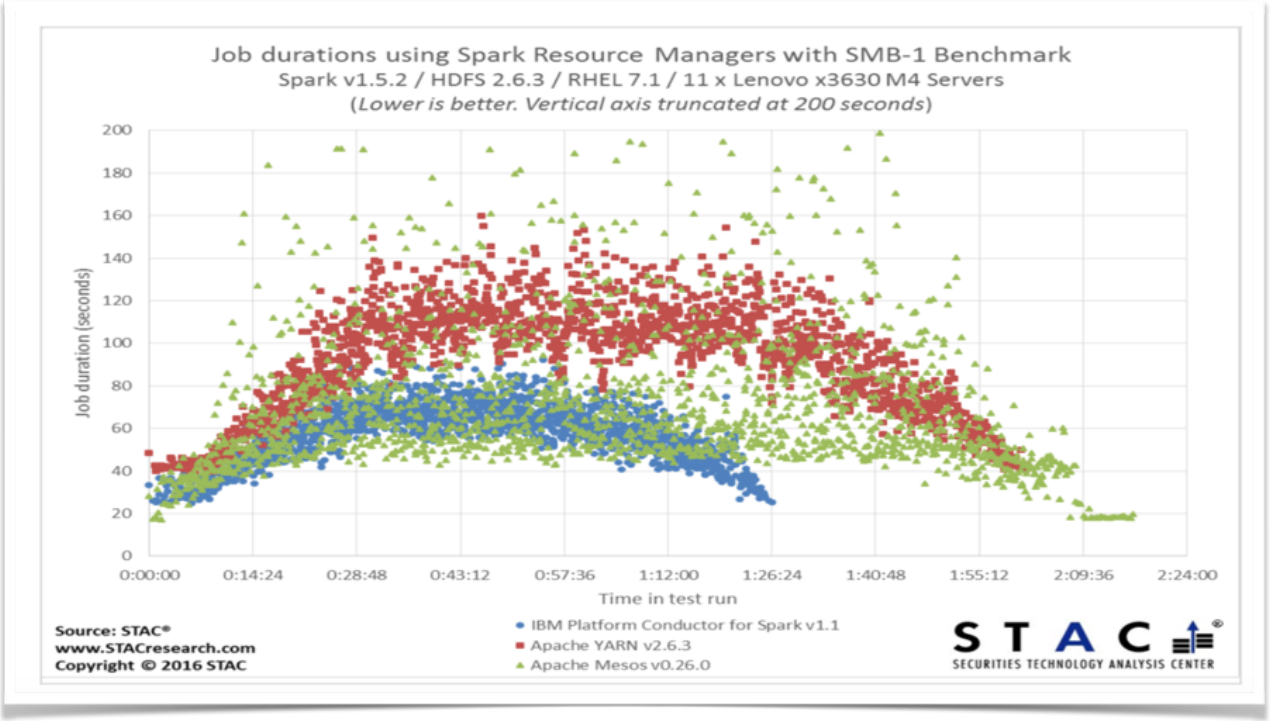


Integration flow

1. Get Pods
2. Match Pods and Offers
3. Bind Pods with Host
4. Update Pods status
5. Run Pods by kubelet



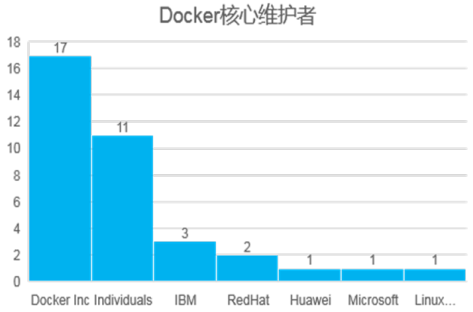
实例：资源调度，资源使用效率



IBM 在社区的贡献

Docker

- 贡献者41位,代码贡献数867次
- 代码贡献数全球第三（仅次于Docker和RedHat）
- IBM在Docker引擎有3位核心维护者
- IBM做出贡献的组件:
  - Docker engine
  - Compose
  - Machine
  - Swarm
  - Distribution (Docker registry)
  - libnetwork
  - libcontainer (prior to OCI)
  - Opencontainers/runC
  - Docker-py
  - User namespace
  - Docker on POWER and z

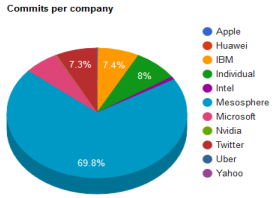


2014年12月4日 Docker与IBM建立战略合作伙伴关系

- DockerHub将支持托管在SoftLayer之上
- IBM将提供WebSphere应用中间件镜像
- Docker容器可以部署在SoftLayer上，无论是裸机还是虚拟机
- IBM将通过集成方案和单独产品两种模式生产和销售Docker Hub企业版。
- IBM将通过集成了Docker编配引擎的IBM BlueMix提供容器服务。

Mesos

- IBM代码贡献仅次于Mesosphere（统计排名中第二位包括全球所有个人贡献者
- 每周和Mesosphere讨论社区的项目状况
- 指导中国的一些公司参与Mesos社区



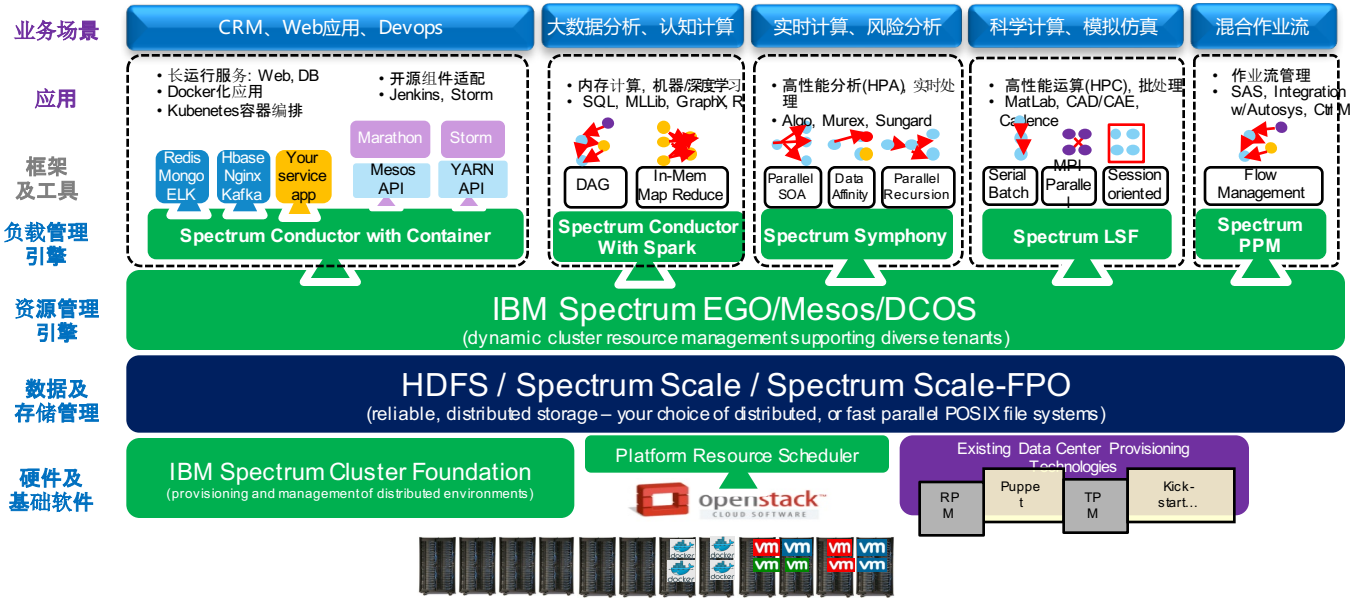
	Company	Commits	LOC
1	Mesosphere	3,357	334,640
2	Individual	490	49,091
3	IBM	403	31,360
4	Twitter	329	37,028
5	Microsoft	283	24,337
6	Intel	34	2,899
7	Apple	32	2,197
8	Huawei	6	81
9	Yahoo	4	357
10	Uber	4	515
11	Nvidia	4	231

Kubernetes

- 2014年7月份加入Kubernetes社区， 主导并贡献了下列的项目：
  - K8s HA cold standby mode (已完成)
  - K8s HA hot standby mode (进行中)
  - K8s ABAC-based authorization that supports changes to the authorization policy without having to restart k8s apiserver (已完成)
  - K8s to support container flavors (进行中)
  - K8s to support POWER(已完成)
  - K8s to support shared quota (e.g., with swarm, spark, etc.) (进行中)
- 超过10个IBM活跃开发者在Kubernetes社区

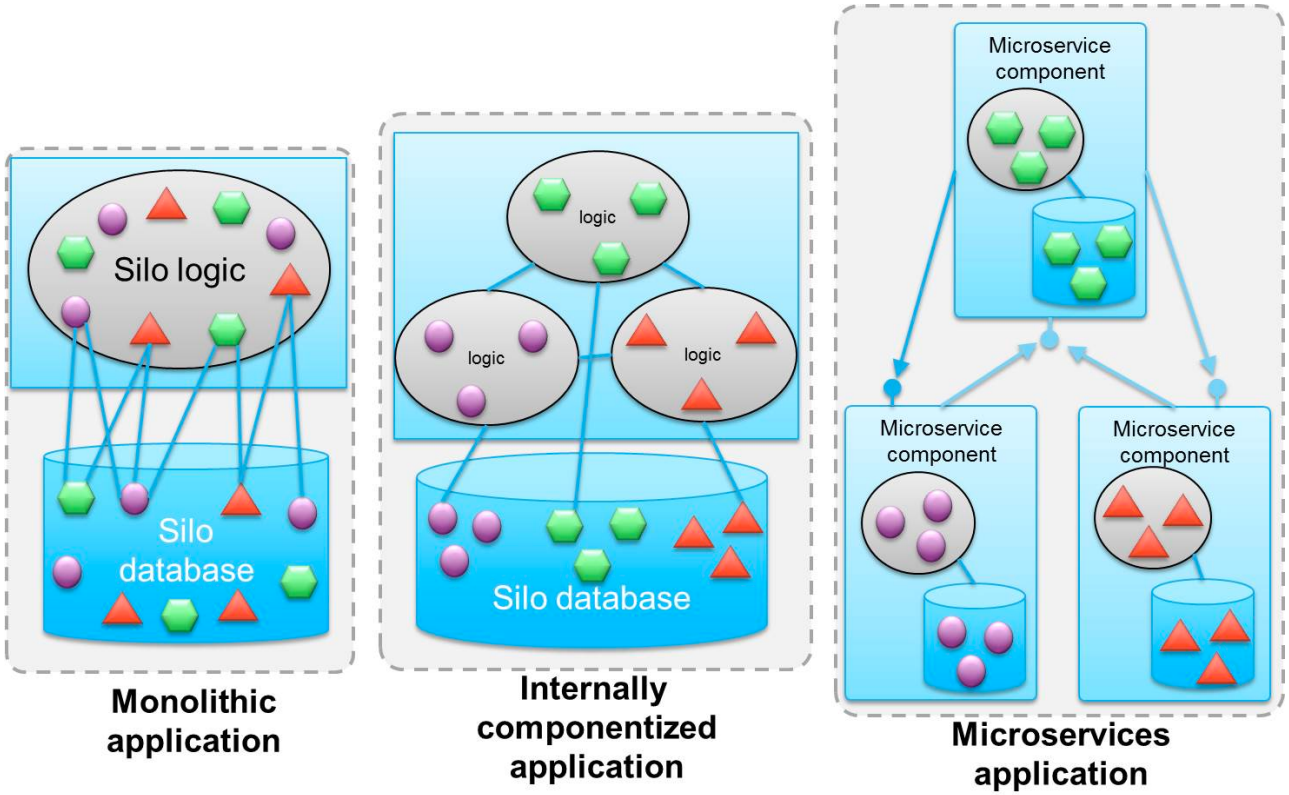
IBM Spectrum 软件定义基础架构平台

支持多种高性能计算，高性能分析，大数据和微服务容器管理框架



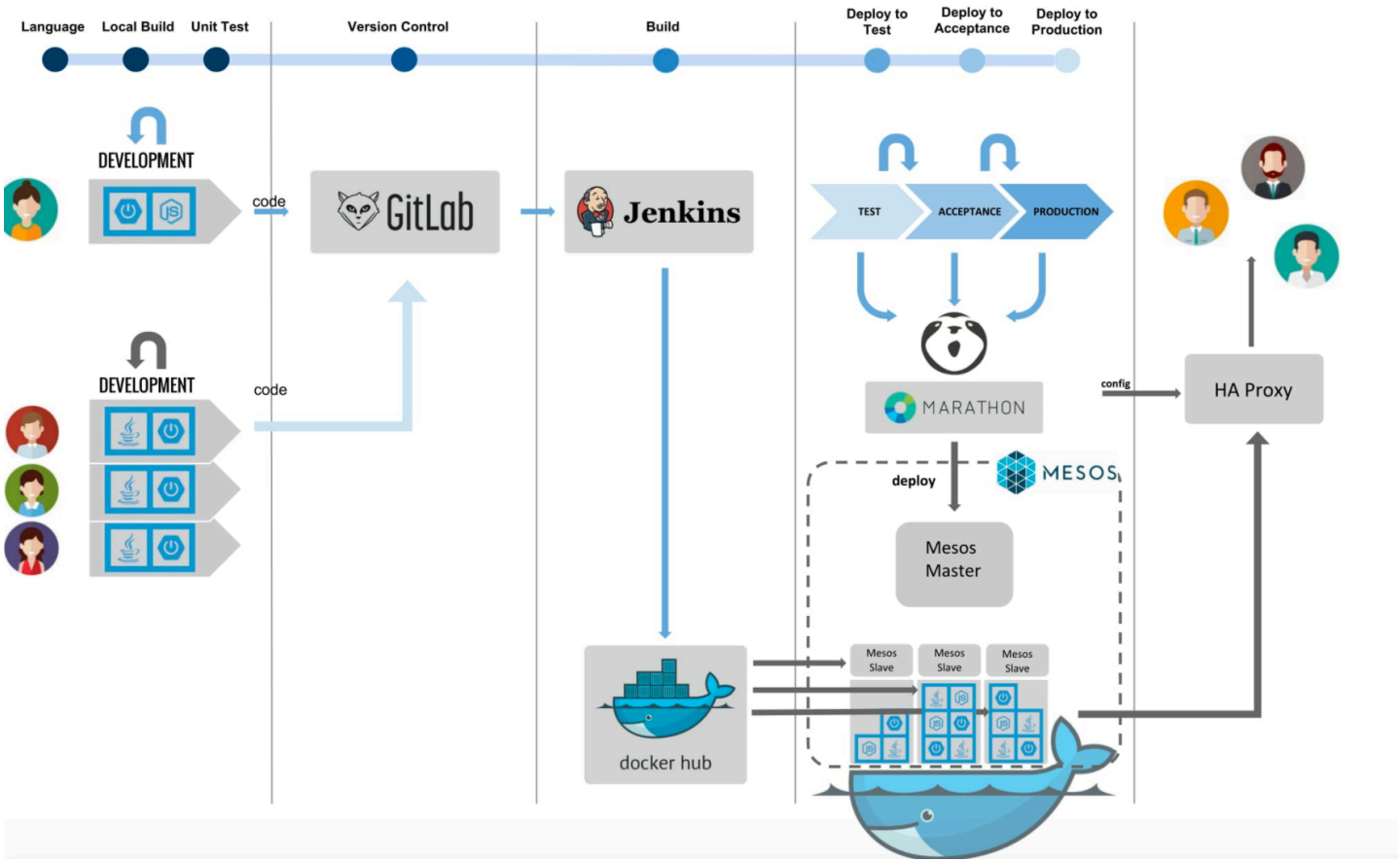
实际生产环境验证的多租户，共享资源框架。支持包括Spark在内的分布式负载。

应用组件架构发展趋势 — 微服务化



- Agility
- Productivity
- Cloud-native
- Scalability
- Resilience
- Stateless

CONTINUOUS DELIVERY WITH MESOS & DOCKER



适用场景

业务流程改进

- 提高 IT 运维效率 e.g. 服务管理，灰色发布
- 软件开发流程、环境改进 (开发，测试和上线) e.g. CI/CD

Docker VS. VM

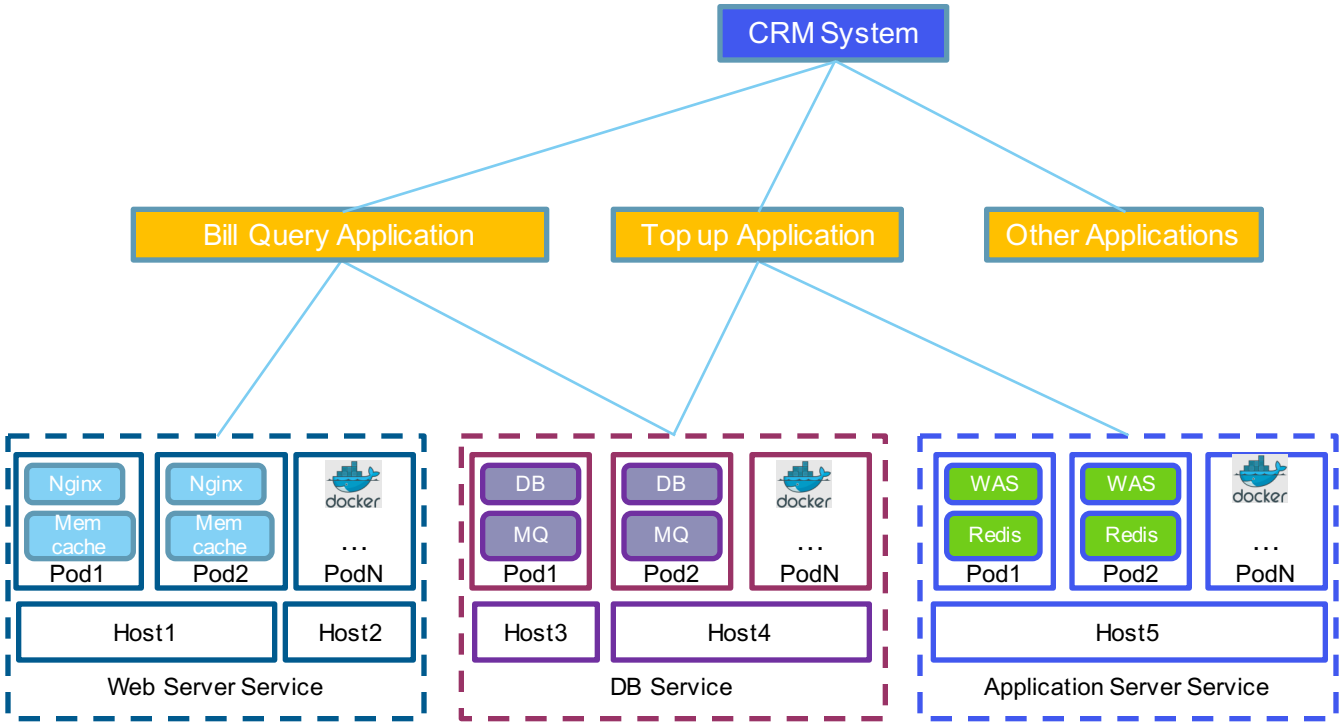
- 运行对性能要求较高的服务 e.g. Billing 系统
- 运行有峰值波动的服务 e.g. 快速启停，抢红包，秒杀
- 提高硬件资源的使用效率 e.g. Docker vs. VM，高级的调度算法 (共享/抢占等)

适用的服务及应用

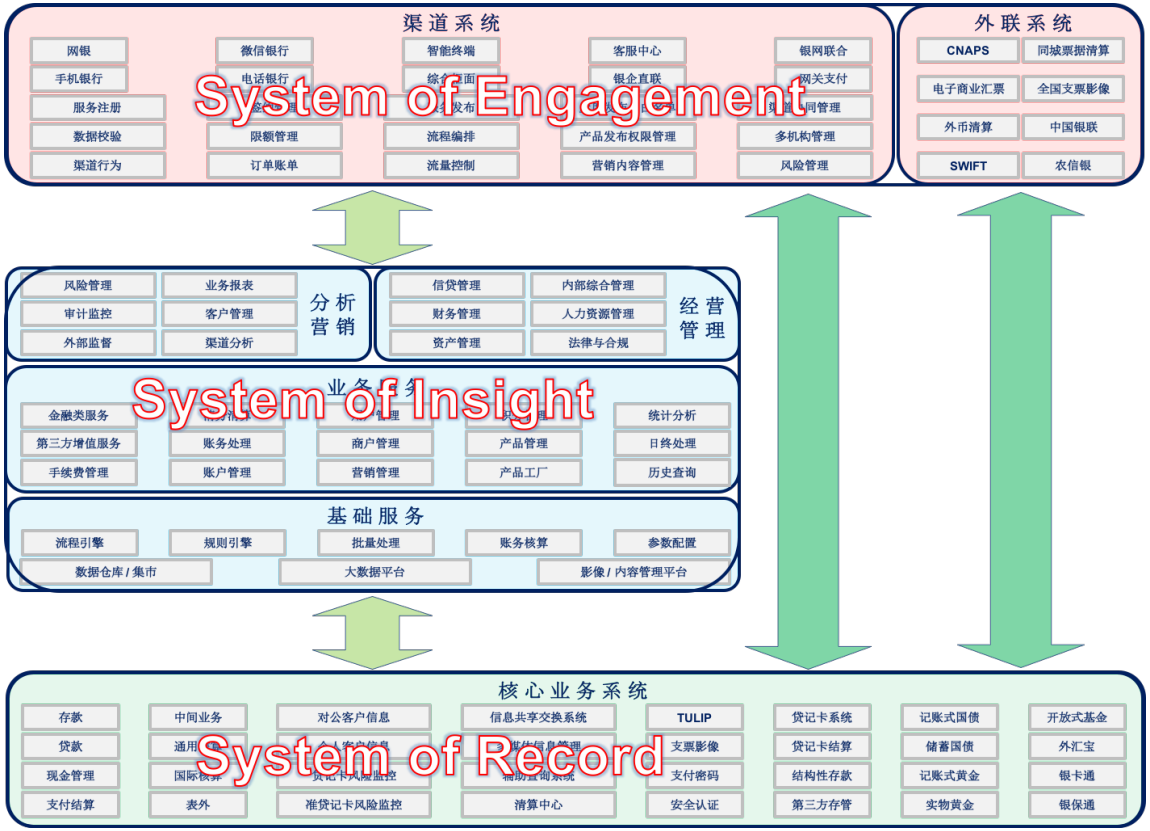
- 无状态服务、应用 – **Yes**
  - Nginx, Tomcat, Memcached
- 有状态服务、应用 – **Yes, case by case**
  - MySQL, HDFS, Cassandra
- 第三方服务、应用 – **Yes, work with vendor**
  - Oracle,

- 不适合的应用
  - 对硬件有特殊要求，e.g. RDMA, GPU

以电信运营商CRM系统为例的业务组件关系



BlueDock/CWC 适用场景 – 银行

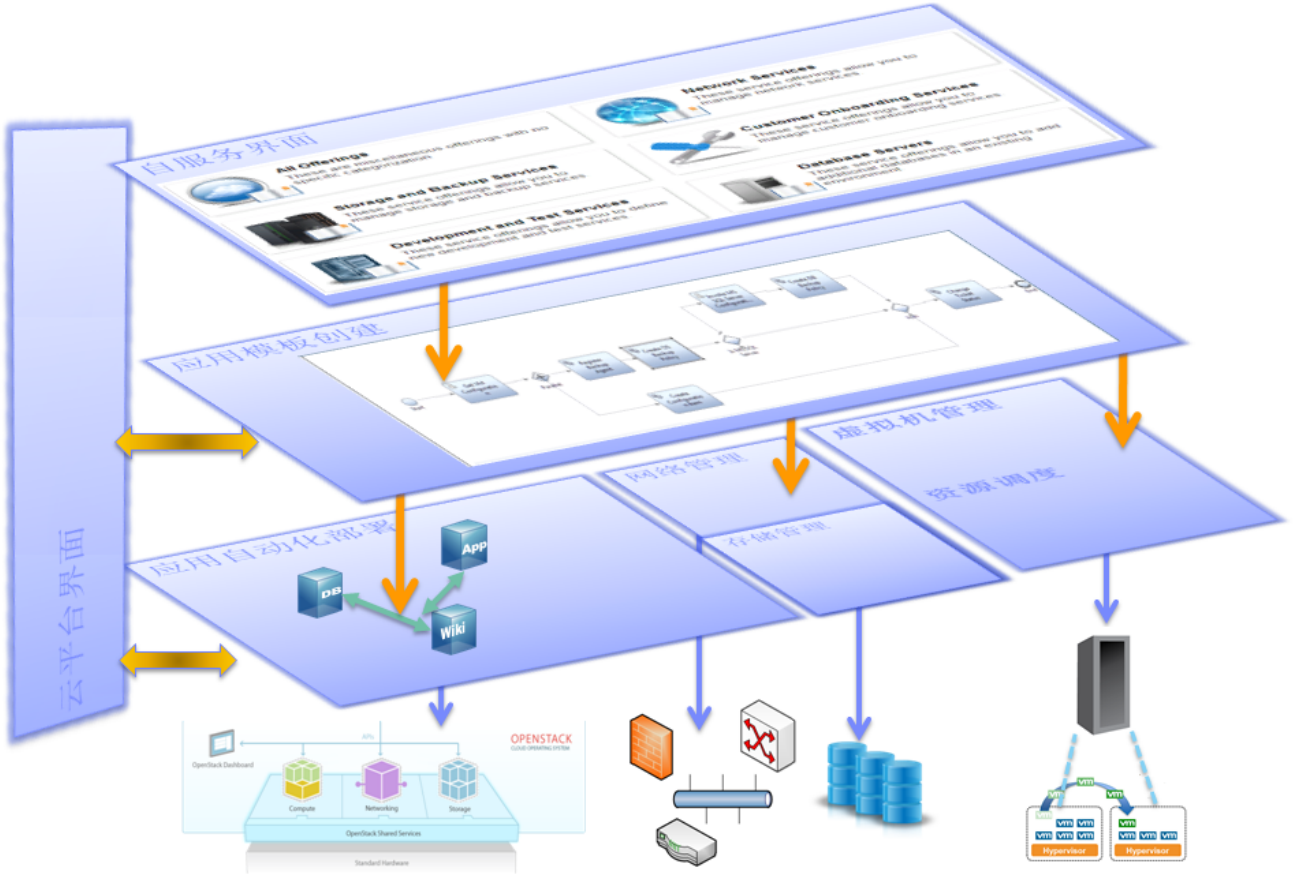


无状态、少状态服务

大数据

数据存储业务

IaaS – OpenStack



PaaS – Cloud Foundry/BlueMix

Compute
- Bare Metal
- Cloud Foundry Runtimes
- OpenStack VMs
- Docker Containers
- Event Driven Apps
- Blueprints (Patterns)
- CMS

Networking
- SDN
- Load Balancer
- VPN

Storage
- Block Storage
- Object Storage

Media
- CDN

Watson
- AlchemyAPI
- Concept Expansion
- Concept Insights, Dialog
- Language Translation
- Natural Language Classifier
- Personality Insights
- Question and Answer
- Relationship Extraction
- Retrieve and Rank
- Speech To Text
- Text to Speech
- Visual Recognition

DevOps
- Active Deploy
- Auto-Scaling
- Delivery Pipeline
- Monitoring and Analytics
- Tracking and Plan GIT
- Image Builder
- Alert Notification

Data & Analytics
- Analytics for Apache Hadoop
- Apache Spark
- BigInsights for Apache Hadoop
- dashDB
- Cloudant NoSQL DB
- DataWorks
- Elasticsearch by Compose
- Geospatial Analytics
- IBM DB2 on Cloud
- Insights for Twitter
- Insights for Weather
- MongoDB by Compose
- PostgreSQL by Compose
- Predictive Analytics
- Redis by Compose
- SQL Database
- Streaming Analytics
- Time Series Database
- Embeddable Reporting

Private APIs
- User Defined Services
- User Defined APIs

Security
- Key Protect
- Security Groups
- IDaaS
- Firewall
- Access Trail
- Application Security Manager
- AppScan Dynamic Analyzer
- Mobile Analyzer for iOS
- AppScan Mobile Analyzer

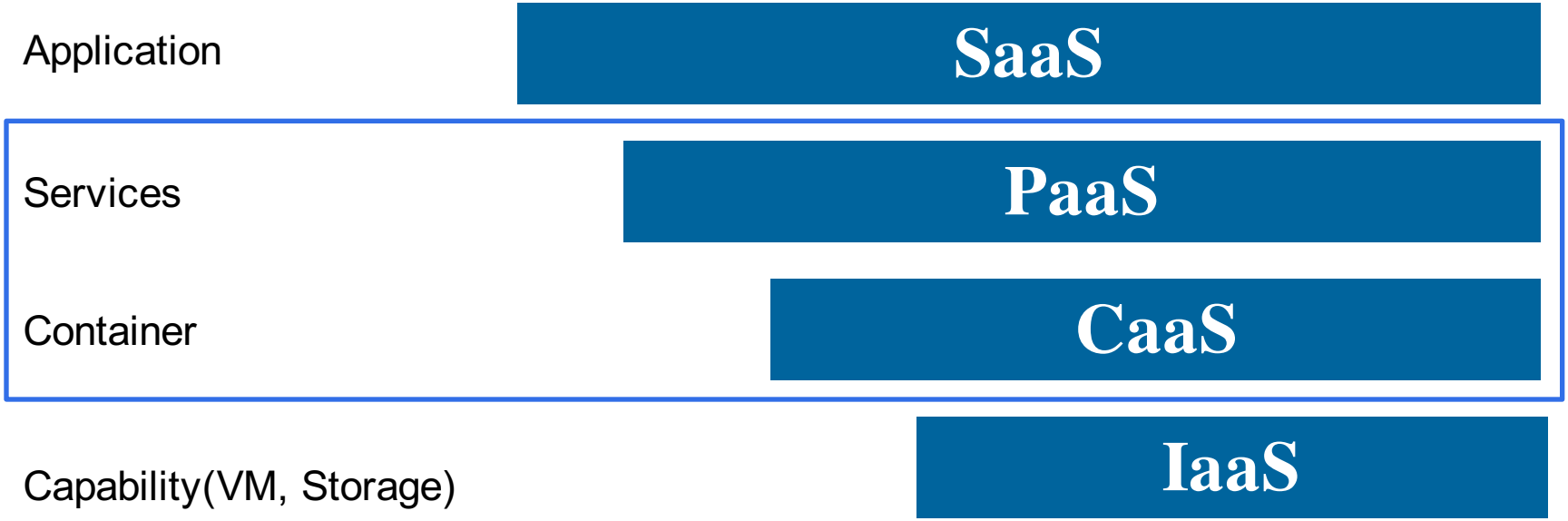
Integrate
- API Management
- Service Discovery
- Secure Gateway
- Service Proxy
- Service Broker
- Cloud Integration

IoT
- IoT Insights
- IoT Real Time Insights

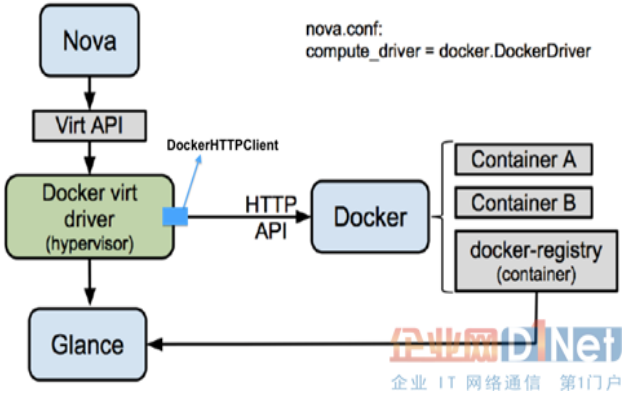
Application
- Workflow
- Big Insights
- Application Server on Cloud
- Business Rules
- Data Cache
- Message Hub
- MQ Light
- Session Cache
- Workflow Scheduler

Mobile
- Presence Insights
- Push
- Mobile Client Access
- Mobile Data
- Quality Assurance
- IBM Push Notifications
- Mobile Application Security

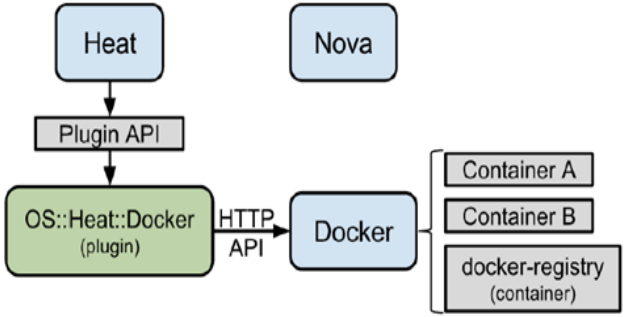
云技术堆栈


Application
SaaS
Services
PaaS
Container
CaaS
Capability(VM, Storage)
IaaS

DCOS与OpenStack集成架构 - 1



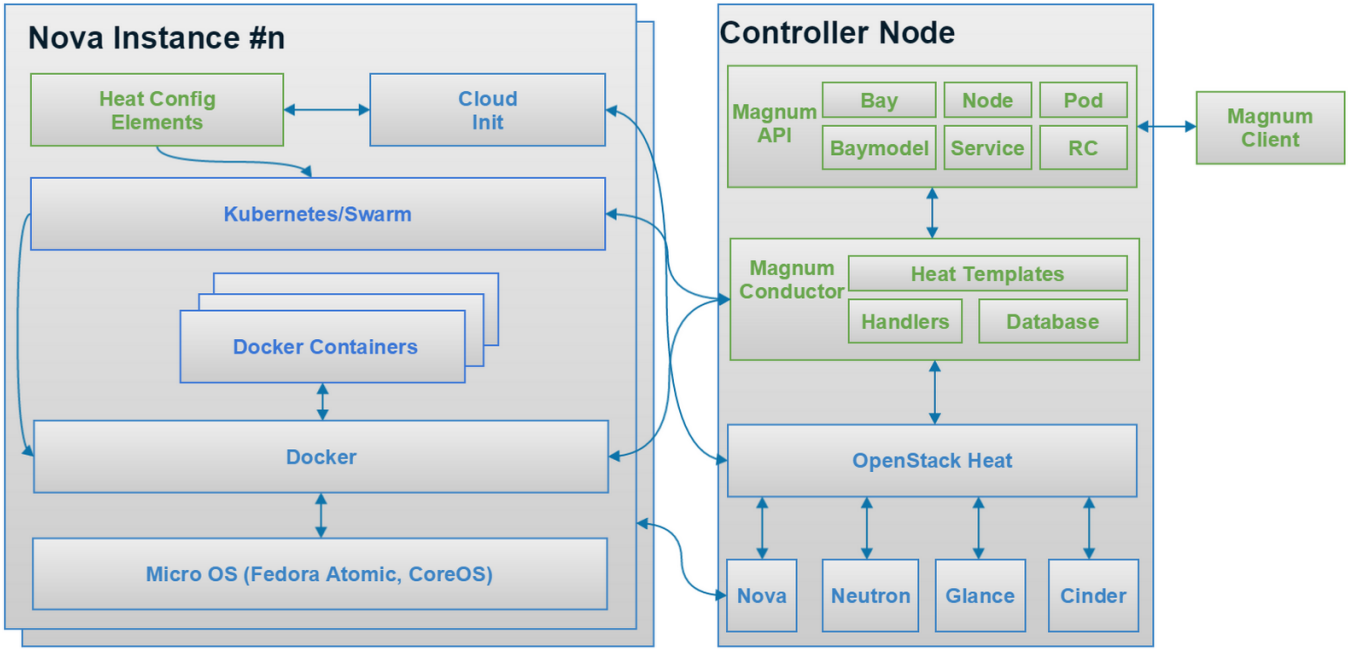
Nova Dcoker Driver



Heat Heat Driver

- Nova Docker Driver
  - Docker作为一种新的Hypervisor, DockerAPI映射成Nova Virt API, Container作为VM, Nova Scheduler调度, Heatheat来做应用部署，服务发现，扩容缩容, Neutron管理Docker网络，支持多租户；
  - 缺点是的高级容器关联、编排、网络功能无法实现
- Heat Docker Driver
  - Heat直接调用Docker API, 因此Heat模版里可以定义所有Docker高级功能
  - 缺点是没有调度，需要指定固定部署服务器，只能使用Docker自带网络，需要手动配置flannel, ovs等网络

DCOS与OpenStack集成架构 - 2



- Magnum项目
  - 支持K8s, Swarm, Mesos的集群（Bay）的定义及创建，支持K8s种的Pod，Service，RC等
  - 使用容器引擎本身的调度和网络

Questions?

http://dss.cn.edst.ibm.com:81/campus/#/login?checkIn=Y&eventId=263
扫码注册&签到；若已经注册，用手机号再次登陆即完成签到。