云计算开源产业联盟

OpenSource Cloud Alliance for industRy,OSCAR

中国云计算开源产业发展白皮书

第三部分 基于 DC/OS 技术的产业

云计算开源产业联盟

Open Source Cloud Alliance for industRy，OSCAR

2017 年 4 月

云计算开源产业联盟

中国云计算开源产业发展白皮书

第三部分 基于 DC/OS 技术的产业

云计算开源产业联盟

Open Source Cloud Alliance for industRy，OSCAR

2017 年 4 月

云计算开源产业联盟

云计算开源产业联盟

目　录

云计算开源产业联盟

文章 I.　版权声明

本调查报告版权属于云计算开源产业联盟，并受法律保护。转载、摘编或利用其它方式使用本调查报告文字或者观点的，应注明“来源：云计算开源产业联盟”。违反上述声明者，本联盟将追究其相关法律责任。

云计算开源产业联盟

文章 II. 前言

随着近几年云计算的爆发式增长，开源技术也在云计算领域得到了新的发展契机，如 OpenStack、Docker、DC/OS 等。云计算开源产业联盟经过深入市场调研，对基于各种开源技术的产业及其各自在中国的市场的发展进行了梳理，分析了发展瓶颈，并对发展方向做出了预测汇总形成白皮书。

《云计算开源产业发展白皮书第三部分：基于 DC/OS 技术的产业》首先阐述了 DC/OS 的发展历程和技术优势，总结了 DC/OS 的使用场景，分析了 DC/OS 在全球，特别是在中国的发展情况，并对 DC/OS 的发展趋势做出了预测。

《云计算开源产业发展白皮书第一部分：基于 OpenStack 技术的产业》、《云计算开源产业发展白皮书第二部分：基于容器技术的产业》已经上传至联盟开源项目 GitHub：https://github.com/opensourcecloud/manual。

云计算开源产业联盟，是在工业和信息化部信息化和软件服务业司的指导下，2016 年 3 月 9 月，由中国信息通信研究院牵头，联合各大云计算开源技术厂商成立的，挂靠中国通信标准化协会的第三方非营利组织，致力于落实政府云计算开源相关扶持政策，推动云计算开源技术产业化落地，引导云计算开源产业有序健康发展，完善云计算开源全产业链生态，探索国内开源运作机制，提升中国在国际开源的影响力。

联盟目前由中国信息通信研究院、华为技术有限公司、北京易捷思达科技发展有限公司、联想（北京）有限公司、国际商业机器（中国）公司、中国电信股份有限公司云计算分公司、中国移动苏州研发中心、联通云数据有限公司、中兴通讯股份有限公司、九州云信息科技有限公司、北京云途腾科技有限责任公司、烽火通信科技股份有限公司、上海优铭云计算有限公司、上海浪潮云计算服务有限公司、杭州华三通信技术有限公司（北京研究所）、杭州云霁科技有限公司、北京奇安信科技有限公司、北京中联润通信息技术有限公司、云栈科技（北京）有限公司、华云数据技术服务有限公司、航天信息股份有限公司、北京云基数技术有限公司（CloudIn 云英）、深圳市深信服电子科技有限公司、上海有孚网络股份有限公司、甲骨文（中国）软件系统有限公司、上海道客网络科技有限公司、上海思华科技股份有限公司、大唐高鸿数据网络技术股份有限公司、上海宽带技术及应用工程研究中心、天津南大通用数据技术股份有限公司、苏州博纳讯动软件有限公司、赛特斯信息科技股份有限公司、国家新闻出版广电总局广播电视规划院、北京国电通网络技术有限公司、携程计算机技术（上海）有限公司、乐视云计算有限公司、中国银联电子商务与电子支付国家工程实验室、北京京东尚科信息技术有限公司 38 家单位组成。

文章 III.　参与编写单位（排名不分先后）

中国信息通信研究院、Mesosphere、烽火通信科技股份有限公司、中国联合网络通信有限公司软件研究院、联通云数据有限公司、北京丝合科技有限公司、华为技术有限公司、国际商业机器（中国）公司、上海浪潮云计算服务有限公司、北京易捷思达科技发展有限公司、联想（北京）有限公司、中国电信股份有限公司云计算分公司、中国移动苏州研发中心、中兴通讯股份有限公司、九州云信息科技有限公司、北京云途腾科技有限责任公司、上海优铭云计算有限公司、杭州华三通信技术有限公司（北京研究所）、杭州云霁科技有限公司、北京奇安信科技有限公司、北京中联润通信息技术有限公司、云栈科技（北京）有限公司、华云数据技术服务有限公司、航天信息股份有限公司、北京云基数技术有限公司（CloudIn 云英）、深圳市深信服电子科技有限公司、上海有孚网络股份有限公司、甲骨文（中国）软件系统有限公司、上海道客网络科技有限公司、上海思华科技股份有限公司、大唐高鸿数据网络技术股份有限公司、上海宽带技术及应用工程研究中心、天津南大通用数据技术股份有限公司、苏州博纳讯动软件有限公司、赛特斯信息科技股份有限公司、国家新闻出版广电总局广播电视规划院、北京国电通网络技术有限公司、携程计算机技术（上海）有限公司、乐视云计算有限公司、中国银联电子商务与电子支付国家工程实验室、北京京东尚科信息技术有限公司

文章 IV.　主要撰稿人（排名不分先后）

陈冉（Mesosphere）、陈文弢（中国信通院）、Ben Lin（Mesosphere）、耿向东（联通）、陈斌（联通）、黄立伟（烽火通信）、罗辉（丝合）、马达（IBM）、郭迎春（IBM）、张俊（华为）、徐伟（浪潮）、何宝宏（中国信通院）、栗蔚（中国信通院）、陈屹力（中国信通院）、马飞（中国信通院）、郭雪（中国信通院）、陈凯（中国信通院）、牛晓玲（中国信通院）、闫丹（中国信通院）

云计算开源产业联盟

一、 DC/OS 发展概述

DC/OS 是一种基于 Apache Mesos 的分布式操作系统，支持对大批量主机的管理，提供自动化资源管理、进程安排、加速进程间交互，并支持简化分布式服务的安装和管理。使用者可通过 WEB 界面和命令行完成对 DC/OS 的管理、集群和服务状态监控。

Apache Mesos 是由 Benjamin Hindman、Andy Konwinski、Matei Zaharia、Ali Ghodsi、Anthony D. Joseph、Randy Katz、Scott Shenker 和 Ion Stoica，于 2009 年在加州大学伯克利分校发起的开源集群管理软件研究项目。随后，核心研发成员 Benjamin Hindman 将 Apache Mesos 引入 Twitter，并使得 Twitter、Facebook、苹果等大型 IT 企业开始陆续打造各自特色的基于 Mesos 的数据中心管理方案。2012 年，围绕 Mesos 开展商业活动的 Mesosphere 公司成立[1]；2013 年，Mesosphere 启动了 DC/OS 项目研究。

2014 年 8 月，Mesosphere 获得了首轮 1 千万美元的融资；同年 12 月 8 日，DC/OS 正式发布。2015 年 1 月，Mesosphere 发布两款 DC/OS 产品，包括免费版和企业版。2015 年 9 月，微软 Azure 宣布与 Mesosphere 合作，以 DC/OS 支撑 Azure 容器服务。2016 年 4 月 19 日，Mesosphere 宣布与微软、HPE、NGINX、Puppet 等公司合作发起开源战略，将 DC/OS 全部版本开源。目前，DC/OS 正在为超过 8 万个数据中心提供服务。

二、 DC/OS 技术现状及优势分析

（一） Mesos 解析

Mesos 是 Apache 下的开源分布式资源管理框架，采用简化的 master/slave（agent）结构设计。master 中的所有元数据均可以通

[1]《DCOS（数据中心操作系统）到底是什么鬼》：http://www.wtoutiao.com/p/1c5cHej.html

过 slave（agent）重构，因此，依托 zookeeper，Mesos 可以快速解决 master 单点故障。

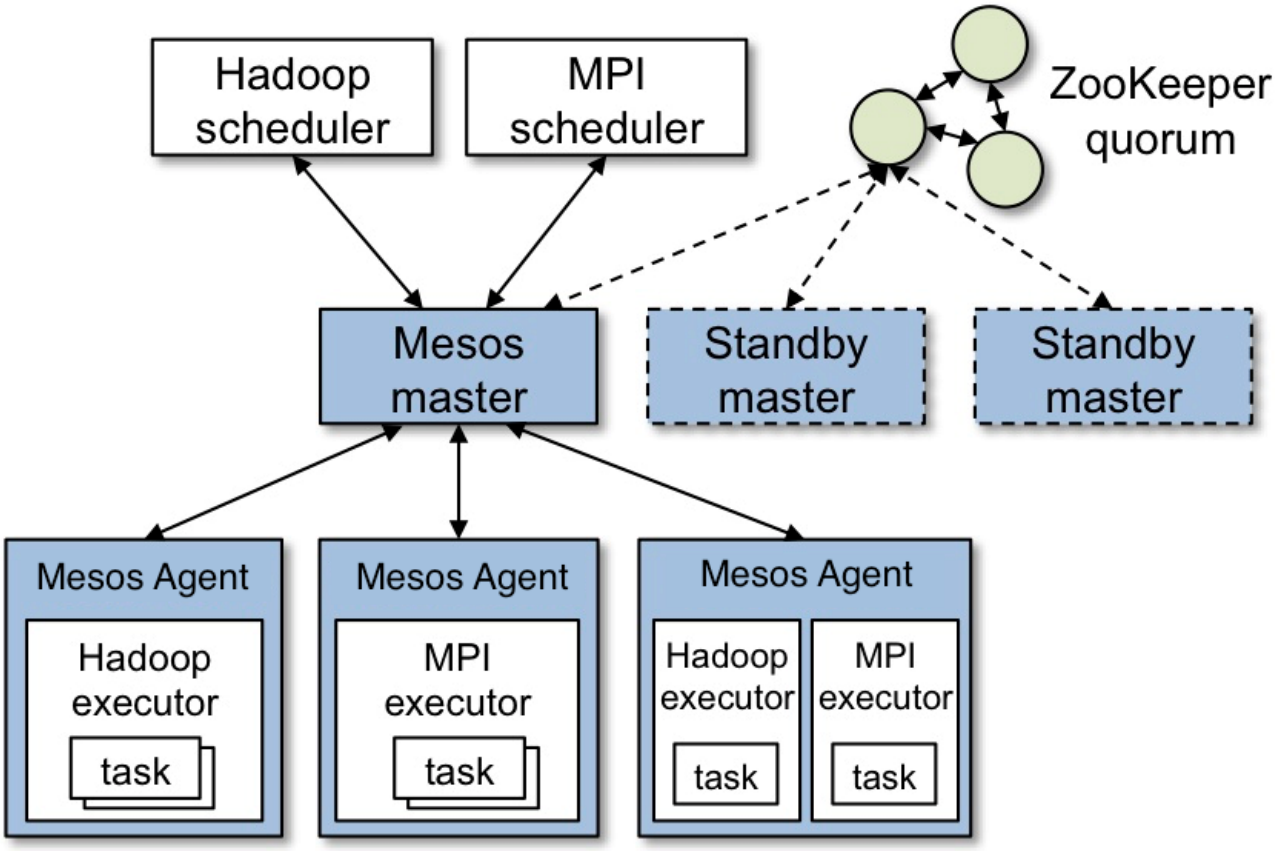

图1 Mesos 架构图（http://mesos.apache.org/documentation/latest/architecture/）

Mesos 由 4 个部分组成：Mesos-master、Mesos-slave（agent）、Framework、Executor。

Mesos-master 是整个系统的核心，负责管理接入 Mesos 的各个 Framework（由 Frameworks manager 管理）和 Slave（由 Slaves manager 管理），并将 Slave 上的资源按照某种策略分配给 Framework（由独立插拔模块 Allocator 管理）。

Mesos-slave 负责接收并执行来自 Mesos-master 的命令、管理节点上的 Mesos-task，并为各个 task 分配资源。Mesos-slave 将自己的资源量发送给 Mesos-master. Mesos-master 中的 Allocator 模块将资源分配给相应的 Framework。

Framework 是指外部的计算框架，如 Hadoop，Mesos 等，这些计算框架可通过注册的方式接入 Mesos，以便 Mesos 进行统一管理和资源分配。Mesos 要求可接入的框架必须有一个调度器模块，该调度器负责框架内部的任务调度。当一个 Framework 想要接入 Mesos 时，需要修改自己的调度器，并向 Mesos 注册,获取分配给自己的资源，再由自己的调度器将这些资源分配给框架中的任务。

Executor 主要用于启动框架内部的 task。由于不同的框架，启动 task 的接口或者方式不同，当一个新的框架要接入 Mesos 时，需要编写一个 Executor，告诉 Mesos 如何启动该框架中的 task。[2]

（二） DC/OS 技术架构

从系统架构上看，DC/OS 架构分为 kernel space 和 user space。其中，kernel space 包括 Mesos Master 和 Mesos Agent；使用区包括集成系统和进程，集成系统包括了 Mesos-DNS、Distributed DNS Proxy，以及 Spark、Marathon 等服务。

[2] Apache Mesos 总体架构：http://dongxicheng.org/apache-mesos/meso-architecture/

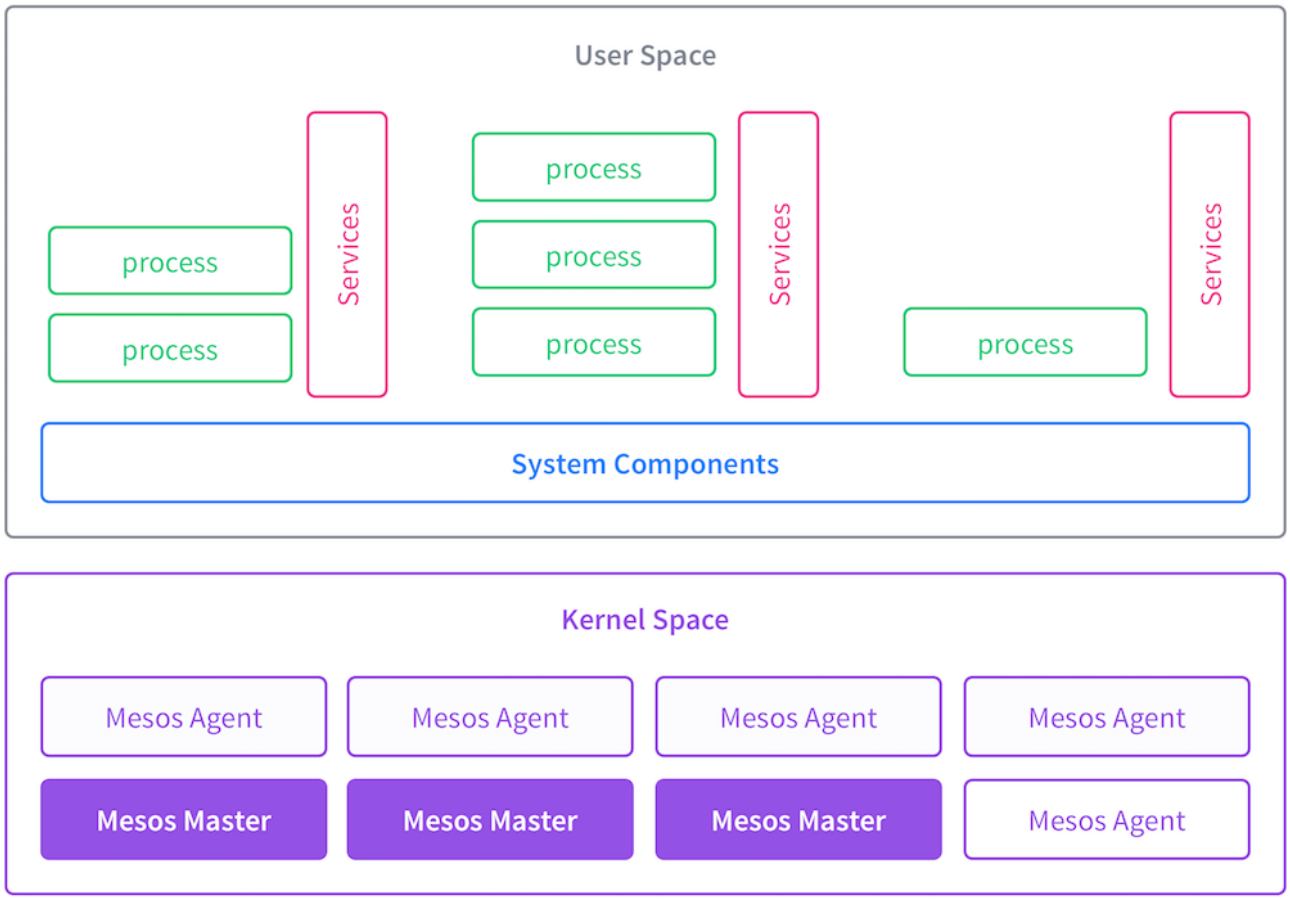

图2 DC/OS 架构图（https://dcos.io/docs/1.8/overview/architecture/）

➢ Kernel space 采取对集群的两级调用完成资源分配。两级调用是通过 Mesos Master 和 Mesos Agent 实现的。

- Mesos Master 负责在代理节点上编排进程，并从 Mesos Agent 上获取资源和进程运行状态，根据运行状态完成对分布式资源的统一调用，用于支撑服务的运行。管理节点通常采取冗余的机制，当主节点不能正常工作时，备节点将自动切换，期间系统保持可用状态。
- Mesos Agent 包括 Private Mesos Agent 和 Public Mesos Agent 两类。其中，Private Mesos Agent 运行用户部署的应用和服务，节点运行在私有网络中，不同私有网络之间相互隔离；Public Mesos Agent 运行 DC/OS 本身的应用和服务，

运行在公网中。Mesos Agent 通过 Mesos-slave 获取本地 CPU、内存资源状态，并将资源状态反馈给 Mesos Master。同时，Mesos Agent 接收来自 Mesos Master 的调用请求，通过 Mesos-containerizers 执行进程任务。Mesos-containerizers 依托 Linux cgroups 和 namespaces 提供了轻量级的容器技术和隔离。

User space 由系统部件和 DC/OS 服务构成。

➢ 系统部件默认集成安装在 DC/OS 集群中，包括：

- Admin Router，基于开源 NGINX 架构，为 DC/OS 服务提供身份鉴别和接入代理。
- Exhibitor，自动化安装部署 Zookeeper，并为其提供 Web UI 界面。
- Mesos-DNS，提供 DC/OS 服务自动发现功能，使不同的应用和服务之间通过 DNS 实现自动发现。
- Distributed DNS Proxy，提供内部 DNS 调度。
- DC/OS Marathon，提供集群管理服务，支持对 DC/OS 应用和服务的启动和监控。
- Zookeeper，提供对 DC/OS 服务的管理和协调。

➢ DC/OS 服务主要由两部分组成，用户应用和服务的编排和执行。

从运行流程看，DC/OS 可分为核心层（Core）、服务层（Service）、应用层（Application）。

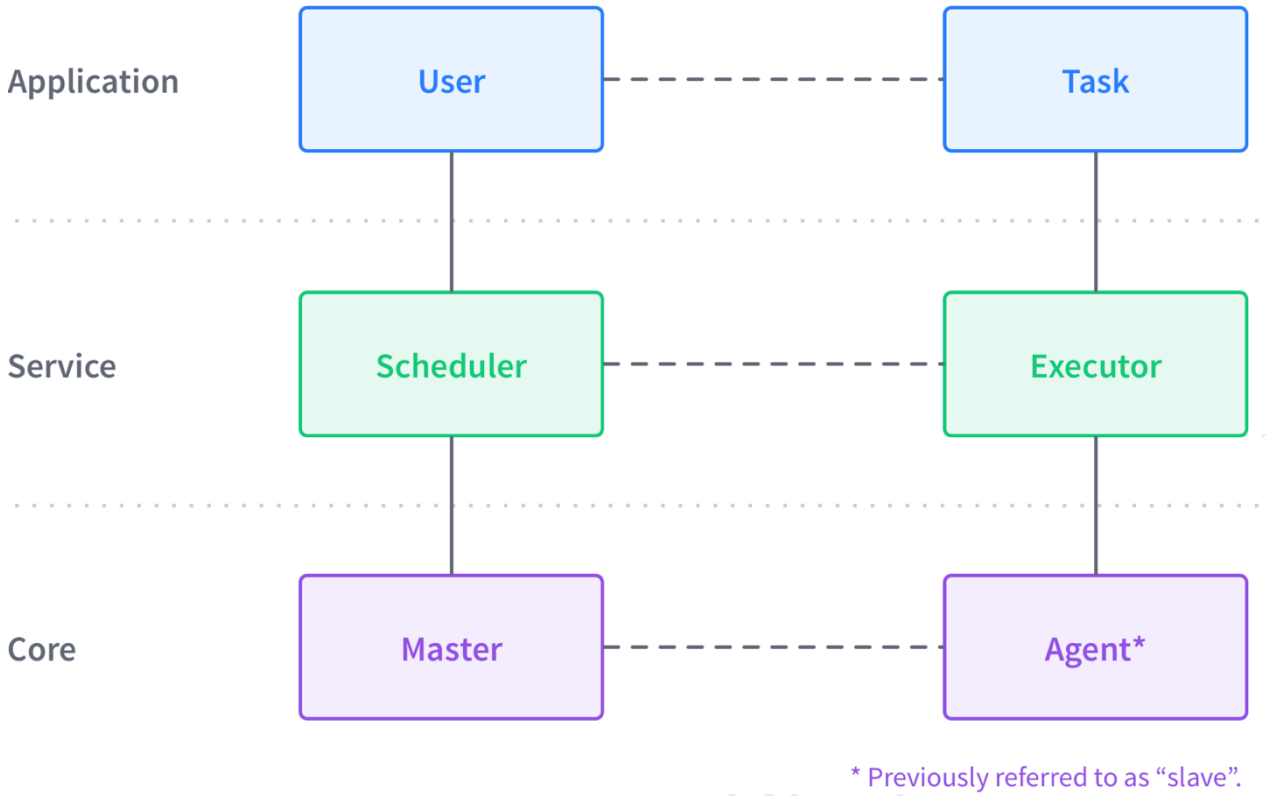

图3 DC/OS 运行架构图（https://dcos.io/docs/1.8/overview/architecture/）

当 DC/OS 进程启动时，DC/OS 的层与层和同一层之间会发生相互的交互：用户首先通过客户端命令行或 Mesos-DNS，向进程调度器（Scheduler）发送进程启动请求。随后，Mesos Master 会依据集群状态和算法，向 Scheduler 分配资源。Scheduler 会根据客户端请求量，逐步释放从 Master 获取的资源，至全部客户端都不再请求进程时，Scheduler 会将全部资源释放回 Master。在 Scheduler 获得了 Master 分配的资源后，客户端即可启动进程；同时，Master 会向 Scheduler 发送资源，当资源足够使用时，Scheduler 向 Master 发送任务启动请求。Master 收到请求后，调度 Agent 通过 Executor 启动进程。进程启动后，DC/OS 中的 Executor、Agent、Master、Scheduler 逐级向客户端报告运行状态。具体运行流程可如下图所示：

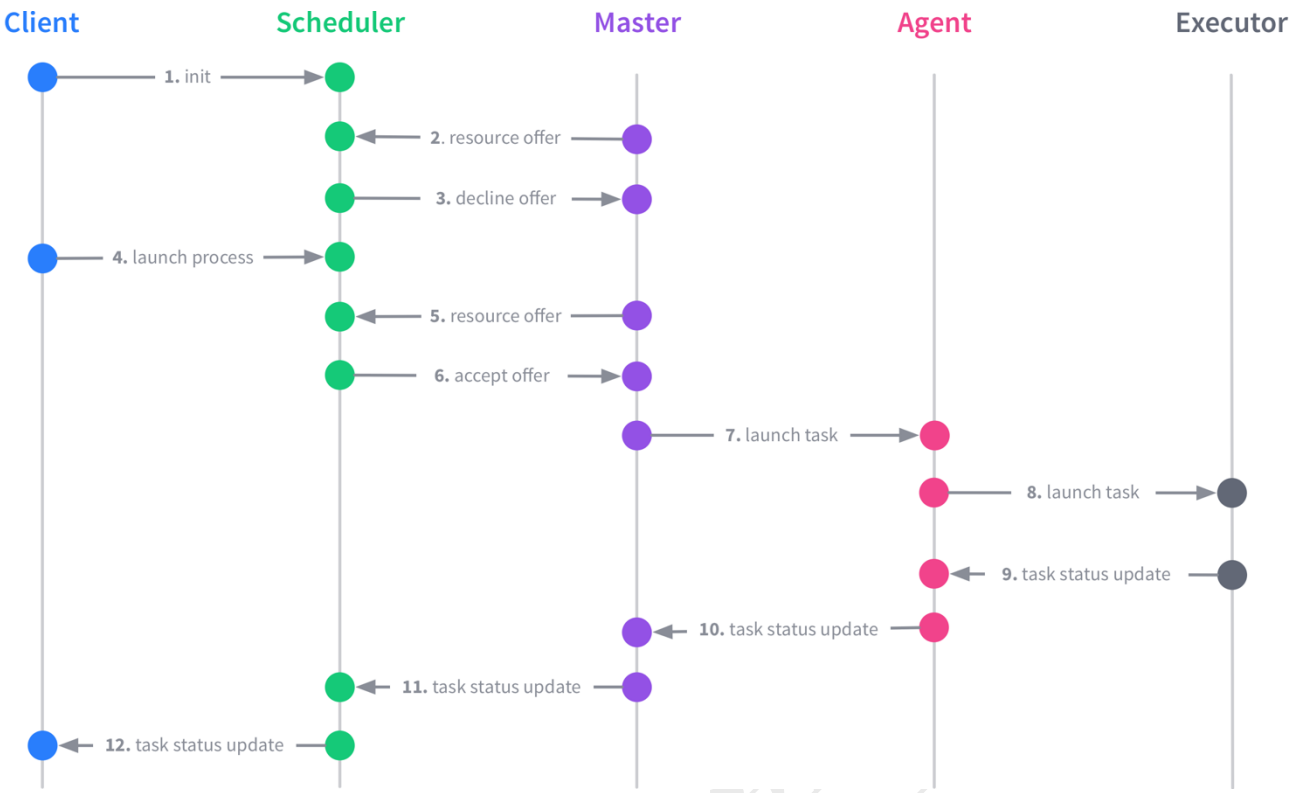

图4 DC/OS 运行流程图（https://dcos.io/docs/1.8/overview/architecture/）

（三） DC/OS 技术优势

➢ 支持快速构建完善的容器系统

由于 DC/OS 基于 Mesos 核心，因此可用于搭建容器和大数据等应用，并将容器和应用以服务的形式运行，同时能够快速迁移至生产环境。DC/OS 可以通过 Marathon 技术，合理编排 Docker 等容器。基于 Mesos 核心构建的 DC/OS，一方面支持灵活的部署容器；另一方面，Mesos 提供了原生的容器工具。原生 Mesos 容器基于 Linux Cgroups 和 Namespaces，提供了容器所需的隔离方法，优势在于用户通过图形化界面或者命令行安装新服务时，Mesos 无需创建镜像，即可自动完成容器的调度和隔离，有效缩短了分布式系统的部署时间。

DC/OS 同时拥有一套独特的双层调度体系，具体如下图 4 所示。

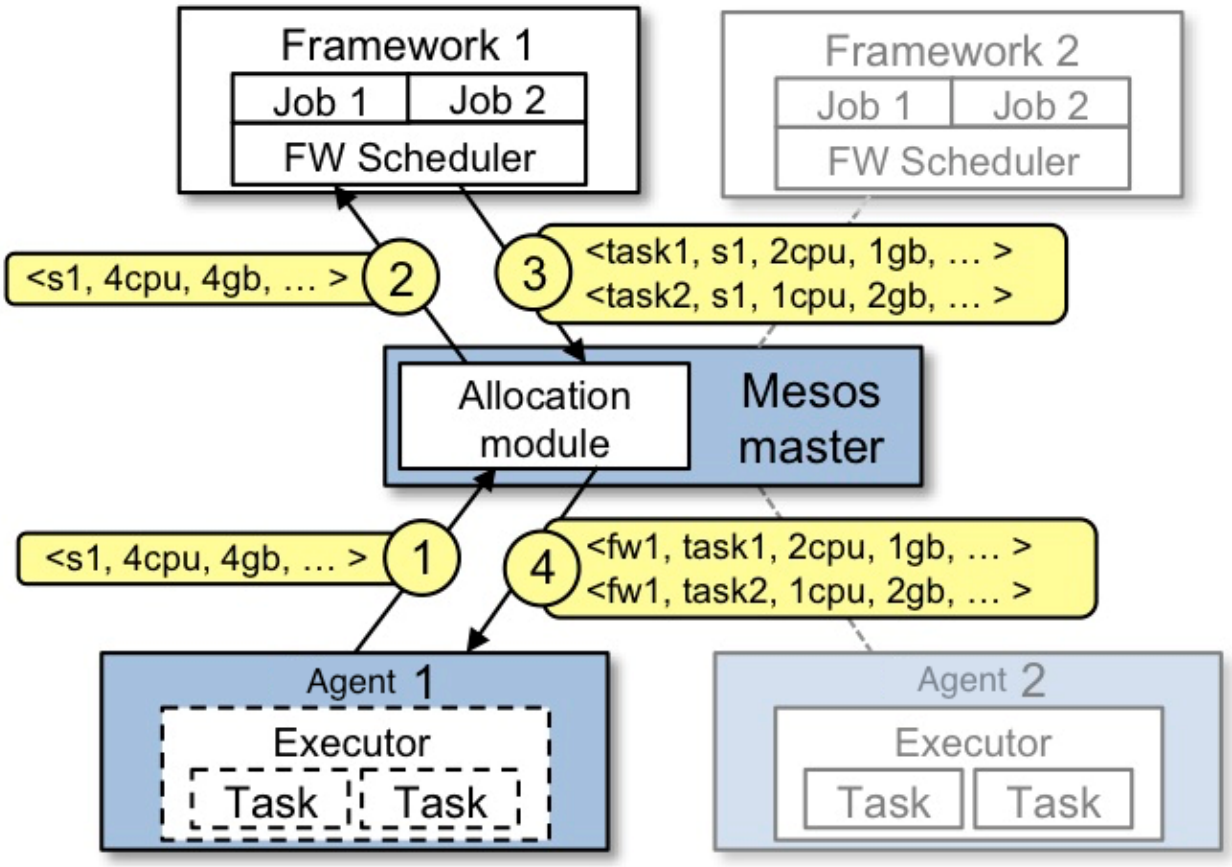

图5 Mesos 任务运行流程图（http://mesos.apache.org/documentation/latest/architecture/）

在 Mesos 中，Framework 和 Mesos master 中均具有 Scheduler 模块（Mesos master 的 Schedule 位于 allocation 中），Mesos master 中的 Scheduler 会将资源按照需求分配给每个 Framework；Framework 中的 Scheduler 会根据资源分配规则，将资源分配给每个任务和服务。Mesos 支持以服务的形式运行容器，能够将 Kubernetes、Swarm 等系统以服务方式运行，体现出了对容器更好地支持。另外，对于 Spark、Kafka 等应用，DC/OS 能够以服务形式，展现在图形界面中。同时，安装和运行在 DC/OS 上的应用和服务全部运行在同一集群内，为客户的管理提供了便利。

DC/OS 提供了 package 管理机制，将应用和服务及所需的配置打包成模板，上传至 package repository，使用者在安装时，只需使用类似 yum 安装方法的一行命令即可完成安装。

以安装 Spark 应用为例，安装流程如下图所示：

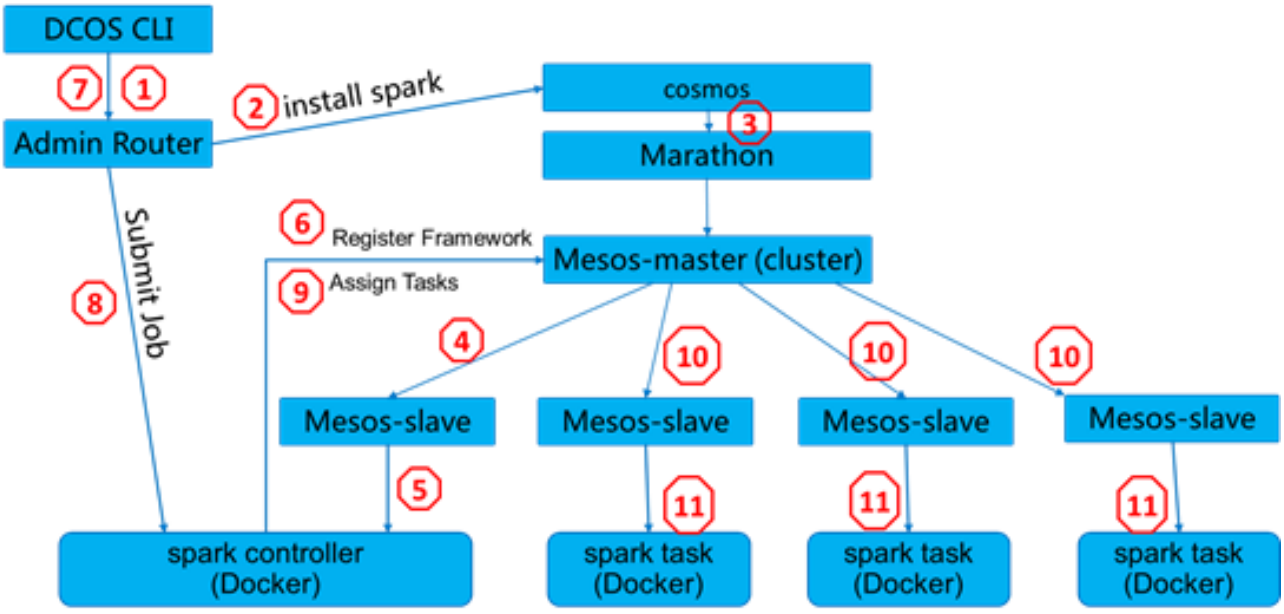

图6 DC/OS 安装应用流程图（http://www.cnblogs.com/popsuper1982/p/5930827.html）

在安装过程中，命令首先通过客户端将请求提交给 Admin Router，随后 Admin Router 将请求通过 cosmos 提交给 Marathon。Marathon 将请求提交给 Mesos master 和 Mesos agent。Mesos agent 启动一个容器运行 spark，并注册到 Mesos 里面成为一个新的 Framework。安装完成后，Spark 被封装在一个 Docker 或者 Mesos 容器中，只有当应用运行时，才会通过 Mesos agent 分配给 Docker 资源支撑应用的运行。与传统的容器技术相比，Mesos 中的应用在运行前不占用资源，传统容器技术中的应用在部署完成时，就已经分配给应用预留资源。

- 支持构建 Spark、Marathon 等服务

Spark 是在 2012 年开源并成为一个 Apache 孵化器项目。Mesos 于 2010 年成为一个 Apache 孵化器项目，这两项技术已经成为成熟的 Apache 顶级项目，迅速完成从学术界到主流企业 IT 商业的跳跃。BenH 在 Twitter 的大规模 Mesos 使用促成了广泛企业对 Mesos 的追捧。BenH 于 2014 年以共同创始人和首席设计师身份加盟 Mesosphere，而 Matei 和 Andy 共同创立基于 Spark 的 Databricks 公司于 2013 年。

此后，Mesos 和 Spark 都专注于如何改变大规模计算中，尽管并行而非彼此并行。Mesos 社区集中在容易地管理共享集群上的分布式系统，而 Spark 社区专注于简化流程并提高实时分析的性能。这两个开源项目现在都非常受欢迎，在世界上一些最大的公司中部署在生产环境中，也应用于最具创新力的公司中。

现在，Mesosphere 正在通过开源 DC/OS 把 Mesos 和 Spark 粘合的更紧密。DC/OS 在 Mesos 周围增加了更多操作的简介性和运维特征，从而解决了在其他平台上很难运行 Spark 问题（包括其他分布式系统）。在 DC/OS 上运行 Spark 的用例和好处包括：

隔离：在大型项目中，团队在不同的部门经常分享一个 Spark 集群，但是可以运行不同的版本。 他们希望在不同版本之间做隔离。并无互相干扰和辅助功能，使每个 Spark 实例访问足够的资源来运行它自己的工作。多版本 Spark 支持在 DC/OS 中是独有的，其他环境中是不可能实现的。

真正的共享资源池基础架构：总体而言，DC/OS 让企业可以运行任何数据库，Web 服务或任何其他 IT 组件在共享资源池中，在 DC/OS 中 Spark 打破数据中心的孤岛问题。与 Spark 使用相同的集群（即使是相同的机器或集群）将使其他服务共同拥有同样的灵活性，并使

Spark 发挥很多功能，特别在数据处理系统需要更紧密地集成在一起时，比如其他现代化应用程序组件、容器和微服务。

性能：DC/OS 还正在优化 DC/OS 和 Spark 的性能最大化，包括确保所有 Spark 任务获得 CPU 和 RAM 等量的新的调度算法。 该功能现在可以在 Spark DC/OS 包中使用，并且将放在 Apache Spark 的 2.0 版本中。

Apache 的 Zepplin：DC/OS 正在整合 Zepplin，Zepplin 是 NFLabs 的一个流行的开源图形项目。Apache Zeppelin 也是一个孵化的 Apache 项目，使用户能够使用 SQL，Scala 等进行数据驱动，交互和协作的文档工具。

所有 Spark 组件一键安装：这包括调度，历史数据，用户界面和其他难以安装的项目组件。使用 Spark History Server，用户可以随时运行和读取事件日志，以重新创建给定作业的结果。

与 Apache Spark 版本紧密互联：DC/OS 的 Spark 的组件和包是最新的并在 24 小时内的。这意味着 DC/OS 用户可以选择在提供运营效率和运行最新版的 Spark 之间进行选择。

天生安全集成 HDFS 集群：用户可以轻松使用内置于 HDFS 的 Kerberos 功能去保护他们的数据，并且很容易与 Spark 进行访问并进行分析。运行时数据通过 SSL 保护。

但是，在构建现代应用程序的数据基础架构时，Spark 只是其中一个难题。 DC/OS 还支持多种用于存储和处理数据的其他技术，包括 Kafka，Cassandra，HDFS 和众多分布式数据库。DC/OS 提供无限的整体解决方案，通过提供 Spark、Kafka、Cassadra、HDFS 作为一个集成的实时大数据平台解决方案。

- 高效可靠的运维功能

DC/OS 通过高效的监控和故障定位工具保证应用和服务的高可用和资源的高利用率。一方面，由于 DC/OS 不仅整合了数据中心中所有物理资源和虚拟资源，还将数据中心的服务发现、负载均衡等统一整合，所以运行在 DC/OS 中的应用和服务使用的是统一的基础设施。在管理和运维运行在大规模数据中心的应用和服务时，DC/OS 提供给用户类似于管理一个单独数据中心的便捷。

另一方面，DC/OS 还具备健康状态监测和恢复的策略。DC/OS 的健康状态监测和恢复包含两个层面：在 DC/OS 平台层面，各类模块和组件具有 2-4 个备用模块，当主节点不可用时，Zookeeper 将自动将其中一个备用模块切换成主模块。在应用和服务层面，Marathon 通过监控 TCP、HTTP 等状态，确认每个应用是否存在不可用的情况，一旦出现某一应用或服务不可用，Marathon 将自动将应用或服务切换至其他副本或直接重启。

同时，DC/OS 还集成了监控和故障修复工具，如 Ngios。在系统层面，DC/OS 通过监控工具对运行在 systemd 中的全部系统组件的状态进行监控；在应用层面，DC/OS 会对正在运行的全部应用和服务的 IP 地址和 DNS 状态进行监控，当出现不可用或网络不通时，DC/OS 会尝试重启这些应用。

➢ 提升企业管理可开发的效率

一方面，DC/OS 可以实现应用的快速交付。DC/OS 提供了自动化交付和持续集成和持续交付（CI/CD）工具，有效缩短应用和软件从测试环境到生产环境的周期。由于 DC/OS 可以运行在裸机、云等环境中，所以 DC/OS 对应用和服务表现出了良好的兼容性，这就使开发人员能够使用 Jenkins、Git、JFrog Artifactory 等工具实现开发和交付；DC/OS 对资源的统一和弹性管理，也帮助开发人员快速部署弹

性、高可用的持续交付平台。此外，DC/OS 的健康状态监测和恢复策略实现了 CI/CD 的高容错。

另一方面，开源的 DC/OS 具备完整的生态环境，DC/OS 的用户会将管理数据中心的流程代码等开源，其他用户可以从 DC/OS 自动获取最新的数据中心操作系统运营和维护相关技术。

三、DC/OS 使用场景

（一） 电信行业

➢ BOSS 微服务架构改造

微服务架构带来的好处可以有效解决目前 BOSS 遇到的问题， 微服务将应用拆分成多个独立的服务后，如何实现服务的注册发现，如何让多个实例服务的配置信息变更及时生效，如何对每个服务进行访问的认证，限流以及负载均衡，如何将分布式部署的服务进行集中管理都是采用微服务架构模式后面临的挑战。着重解决 API 网关，服务注册发现，负载均衡，弹性扩缩容，服务熔断，限流降级，灰度发布，调用链分析等服务管理难题。

➢ 服务熔断降级机制

开源的微服务管理组件，使用 Netflix Hystrix 实现服务的熔断降级，有效隔离故障，降低整体业务故障的风险，防止服务雪崩效应。

➢ API 网关

通过引入 API 网关组件，将微服务的管理集中化，解决了由于服务拆分带来的服务分散难于管理的问题，实现服务统一入口，提供客户端管理，服务调用量统计，服务限流，ACL 规则，认证等功能，做到一点管理。

➢ 微服务的持续集成与交付

无自动化不微服务，自动化包括测试和部署，单一进程的传统应用被拆分为一系列的多进程服务后，意味着开发，调试，测试，监控和部署的复杂度都会增大，必须要有合适的自动化基础设施来支持微服务架构模式。而容器快速启动和镜像仓库天生是为 CI/CD 设计的，以前我们启动一个虚拟机需要几分钟，而启动容器只需要几秒钟，有了这种能力以后集群式的和并行的持续集成与交付才能成为可能。我们在工作实践中引入持续集成与持续交付并引入容器化技术，提供了一个从开发到上线都一致的环境，极大提高软件开发效率并保障软件开发质量。持续交付将微服务镜像提交给 DC/OS 的组件仓库中形成安装部署能力。

- 组件仓库

DC/OS 的组件仓库提供了微服务以及公共组件的安装部署能力。基于容器化的方式，DC/OS 把常用的组件整合到平台里面，提供便捷的按需求部署，比起以往人为手工构建，以小时为单位进行应用部署，现在只要一键安装，几分钟就可以构建需要的组件及环境，满足业务应用的需求，节省了大量部署的工作量，提高部署效率。通过对现有 BOSS 微服务改造后，上架部分 BOSS 应用到组件市场。

- 整体业务情况监控

利用 API 网关的集中管理能力，采集 API 网关调用的日志，以服务 API 的维度统计微服务平台的访问量以及时延，出错等数据，更加直观的反映整个微服务平台的健康状态，以及每个 API 的服务质量，实现一点看全。

- 业务系统与大数据平台集成

通过对业务系统微服务改造，利用 DC/OS 自身大数据一键式分布式部署，从而减少数据传输和传递过程。HDFS、Spark、Kafka、

Hadoop 等的智能无缝接入，能够以最短路径分析业务系统产出的各种数据，从而帮助业务系统微服务业务系统持续改进。

（二） DevOps

DC/OS 通过集成 Jenkins，SVN 和 Gitlab 等版本管理工具实现 DevOps 场景，开发人员通过提交代码到 Gitlab 中，代码会通过审批进入 Jenkins 中进行单元测试，功能测试，集成测试，回归测试等，最终发布应用到 DC/OS 平台，通过对 Jenkins 中的 Docker 插件／管道部署实现自动镜像产生并发布成 DC/OS 组件或者 Marathon 直接部署。同时支持蓝绿发布、灰度发布、紧急修复、版本回滚等发布和部署模式。

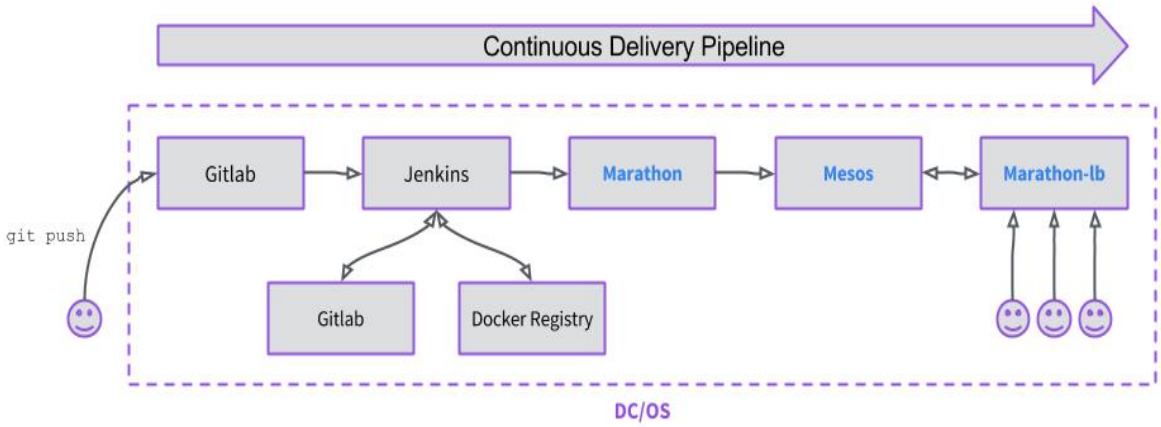

图7 DC/OS 在 DevOps 中的应用（资料来源：Mesosphere）

（三） 弹性扩容（Twitter、eBay）

通过对 DC/OS Marathon 配置实现动态弹性。如下图:

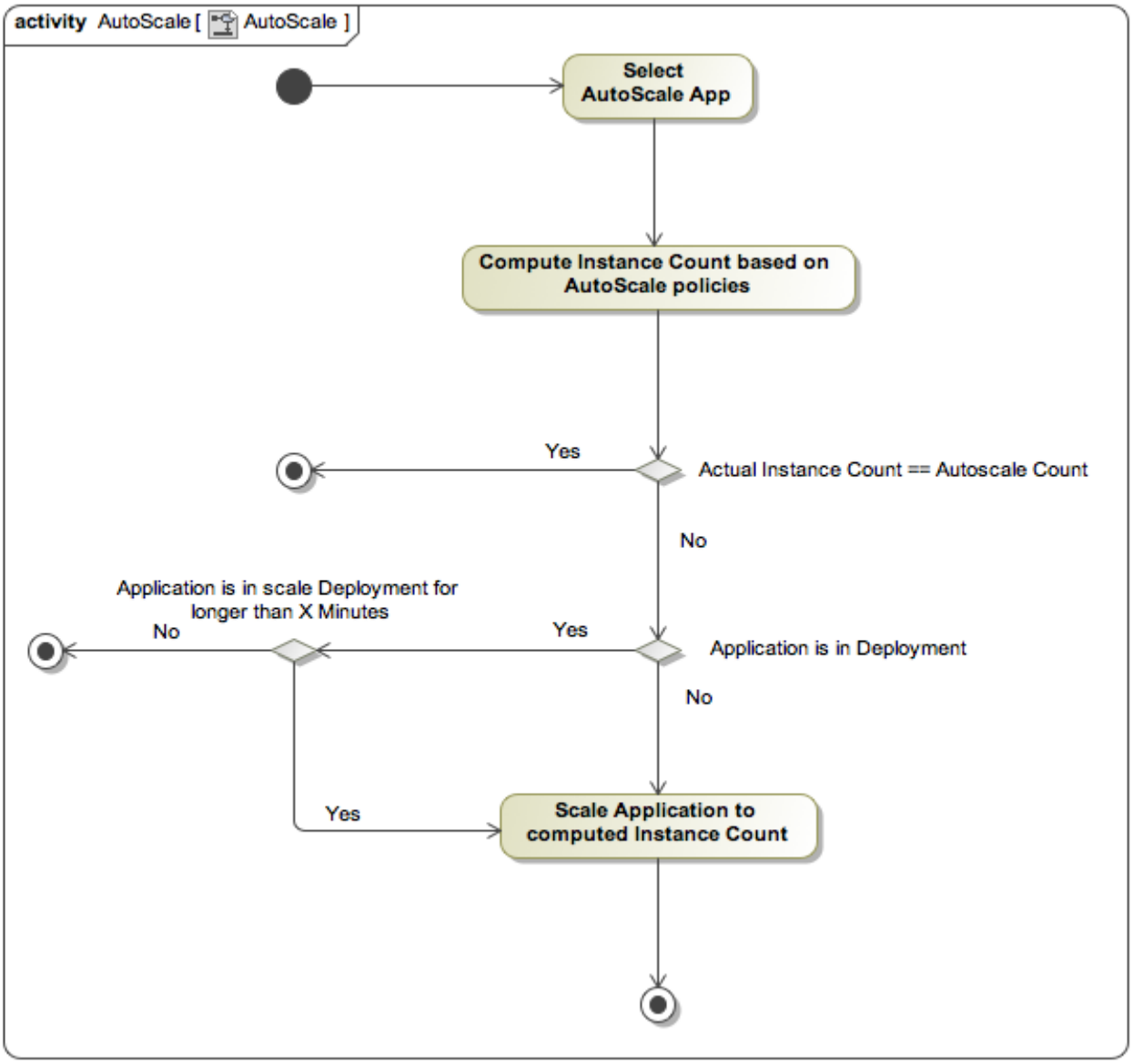

图8 通过 DC/OS 对 Marathon 进行态性调整（资料来源：Mesosphere）

（四） 对裸机应用的管理

已经投产使用的数据中心，其应用部署架构仍是集群化的、计算资源孤岛式的，无法实现应用的自动化部署、扩缩容，因此需要一种较有限的集群管理工具，并可以实现隔离、灰度升级、打包、资源弹性分配等能力。选择 DC/OS 技术可以管理 Docker 容器包装的应用及

服务器集群的 Mesos 技术重构其基础设施架构，建立一种以 Linux 为核心，由遍布数据中心的普通服务器构成的集群。

如果数据中心已经拥有 Hadoop、Spark 等裸机部署的应用系统，而客户又不愿将其容器化，那么 Mesos 技术比 Kuberntes 技术更有优势。因为 Kuberntes 目前只支持容器化应用，对于裸机上的应用无能为力。也就是说当你拥有很多的物理资源并想构建一个巨大的静态的计算集群的时候，Mesos 就派上用场了。有很多的现代化可扩展性的数据处理应用都可以在 Mesos 上运行，包括 Hadoop、Kafka、Spark 等，同时你可以通过容器技术将所有的数据处理应用都运行在一个基础的资源池中。在某个方面来看，Mesos 是一个比 Kubernetes 更加重量级的项目，但是得益于那些像 Mesosphere 一样的贡献者，Mesos 正在变得更加简单并且容易管理。

在 DC/OS 新技术后，在硬件资源和应用部署两方面均有有效的提升，其中硬件资源利用率可以提高 5 以上倍，应用部署效率可以提高 3 倍以上。

四、DC/OS 全球发展现状

（一） DC/OS 全球市场规模及前景

Mesos 正在推动采用现代应用程序方式——容器编排和大数据服务，这同时也是 Mesos 使用的关键驱动因素。虽然 Mesos 被设计用于大规模的操作，但是许多当前用户正在运行具有少于 100 个节点的集群。使用在 Twitter，Uber，Ebay 和 Netflix 进行战斗测试的相同架构，确保客户随着工作量的增加，其集群将无缝扩展。

基于数据调研，Mesos 社区正在快速增长。大多数新公司正在运行 Mesos，将其用于由内部部署，混合和云环境部署的微服务组成的

现代应用程序。用户已经在 DC/OS 提供的功能上看到了巨大的价值，根据市场调研数据：绝大多数新的 Mesos 用户都使用 DC/OS。

DC/OS Universe 是一个类似 App Store 的体验，其开放的目录可以在 DC / OS 上运行。这样，开发人员和运营商就可以轻松安装和运行 Spark，Cassandra，Kafka 和 NGINX 等复杂的分布式系统。DC/OS Universe 简化了其服务的包装和交付，用于数据中心和云环境，为 ISV 提供了低摩擦分配路径，类似于桌面设备和移动设备上的应用商店。 DC/OS 使合作伙伴和个人软件开发人员能够构建在 DC / OS 上运行并将其发布到 Universe 的服务。

（二） DC/OS 社区发展情况

DC/OS 社区

DC/OS 是一个开源项目，于 2016 年 4 月推出，拥有超过 60 多个创始成员，包括埃森哲，思科，惠普企业,微软，联通和 H3C。这些成员承诺通过技术集成，生产部署和开源软件贡献来帮助扩大和塑造项目。DC/OS 社区（https://dcos.io/community）由 Mesosphere 公司发起，思科、NGINX、VERIZON 等公司支持，主要经营项目为 DC/OS，项目在 GitHub 上已经得到超过 1 万次 Commits。从 2013 年 7 月起，DC/OS 项目开始从 GitHub 上逐步迁移，并托管在 JIRA（http://jira.mesosphere.com）上，贡献者可以通过创建 Issue 的方式提交贡献，贡献形式主要包括 Bug、Epic、Improvement、New Feature、Story、Suggestion、Task、Subtask8 类。截止 2017 年 4 月，JIRA 已经收到 Issues 6788 条，其中 Task 类 5723 条，Bug 类 592 条。

在 2016 年，社区用户在全球组织超过 52 个 Mesos 聚会，并在美国，欧洲和亚洲组织三个 MesosCons。在 MesosCon 发言的公司包括

Yelp，Verizon，IBM，Netflix，Intel，华为，Uber，PayPal 和 Twitter。过去一年，Mesos 贡献者的数量翻了一番。在中国地区的活动方面，2016 年 5 月，Linker networks 和 Mesosphere 共同在北京召开发布会，宣布正式在大中华区发布 DC/OS 产品；2016 年 11 月，DC/OS 社区在杭州举办 Mesos con 亚洲大会，分享 Mesos、DC/OS 项目及不断发展的生态系统。社区还拟定于 2017 年在北京等地继续举办大会，推广 DC/OS 技术和生态。

云计算开源产业联盟

云计算开源产业联盟成立于 2016 年 3 月，由中国信息通信研究院牵头，联合各大云计算开源技术厂商，共同在工业和信息化部信息化和软件服务业司的指导下，致力于落实政府云计算开源相关扶持政策，推动云计算开源技术产业化落地，引导云计算开源产业有序健康发展，完善云计算开源全产业链生态，探索国内开源运作机制，提升中国在国际开源的影响力。2016 年 7 月，云计算开源产业联盟发布开源项目，设立 DC/OS 工作组，邀请 DCOS 中国社区创始人、Linker Networks 首席技术官和技术副总裁陈冉担任组长，着力推动 DC/OS 在中国云计算市场的影响和技术落地。

五、 DC/OS 在中国发展情况

（一） DC/OS 在中国市场的发展才刚刚起步

2017 年，DC/OS 中国社区发布对中国云计算企业的调研访问报告《2017 年中国 Mesos、DC/OS 调研报告》，报告显示虽然中国市场对 Mesos 有普遍的了解，但在生产环境中的使用率仍然很低，仅占 10%。有 7%的受访者表示从未听说过 Mesos；有 83%的受访者了解 Mesos，但没有投入使用。

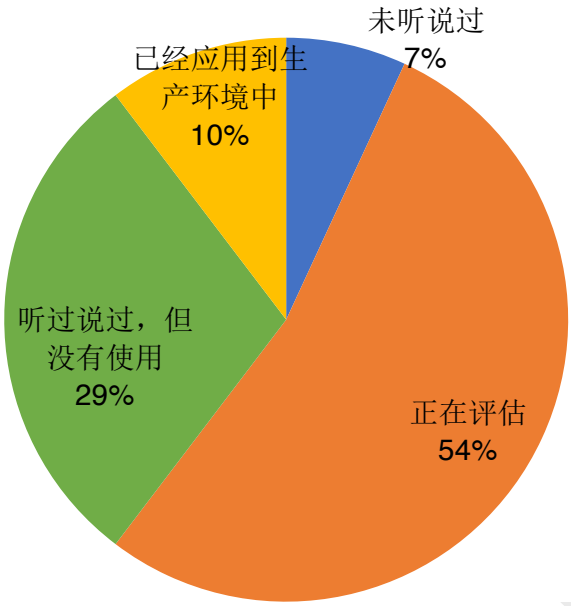

图9　中国市场对 Mesos 的了解程度（资料来源：2017 年中国 Mesos、DC/OS 调研报告）

从行业分布上看，在使用 DC/OS 的企业中，电信、影视、移动互联、服务业占比最高。

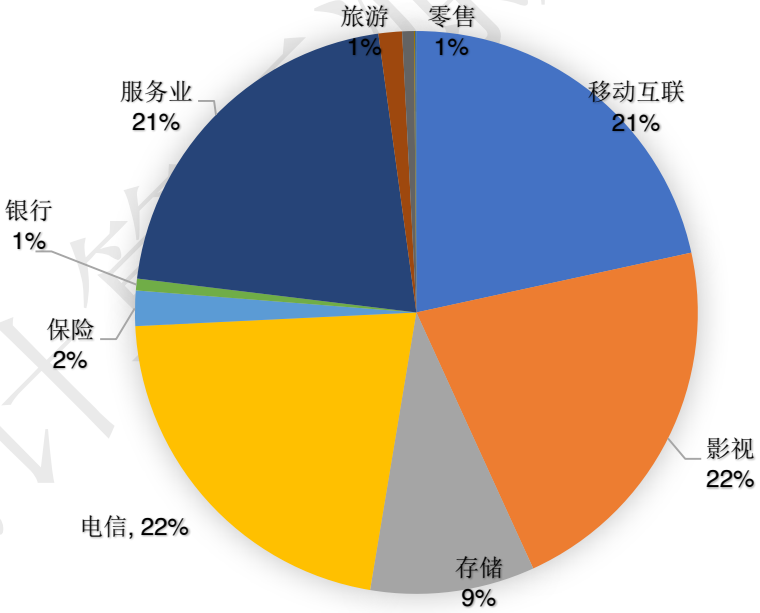

图10 DC/OS 用户行业分布（资料来源：2017 年中国 Mesos、DC/OS 调研报告）

在组件中，Marathon、Kafka、k8s 是使用率最高的组件。

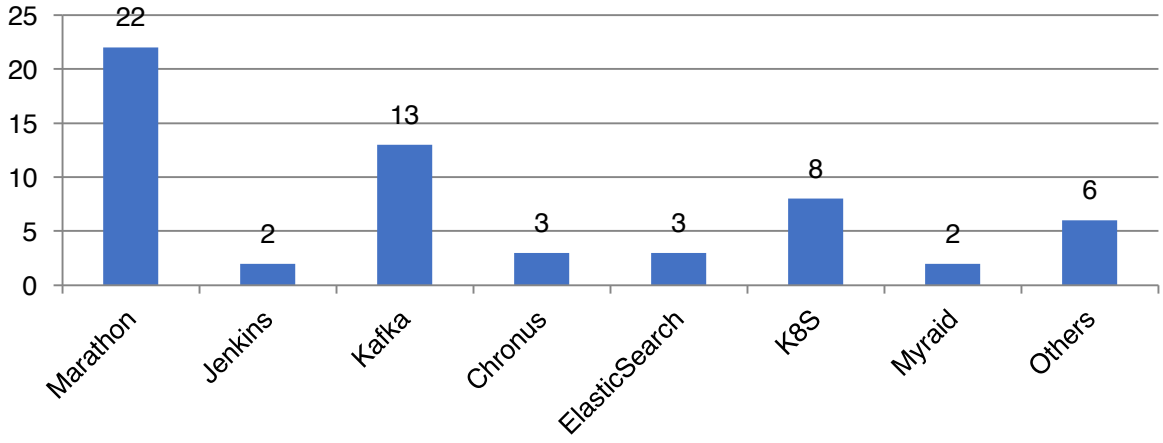

图11 DC/OS 组件使用率（资料来源：2017 年中国 Mesos、DC/OS 调研报告）

在已经使用 Mesos、DC/OS 的用户中，Mesos、DC/OS 提供的稳定和成熟的技术是选用的最主要原因，此外，有效提高资源使用效率和减少开支、提供简单便捷的运维管理功能也是选用的主要原因。

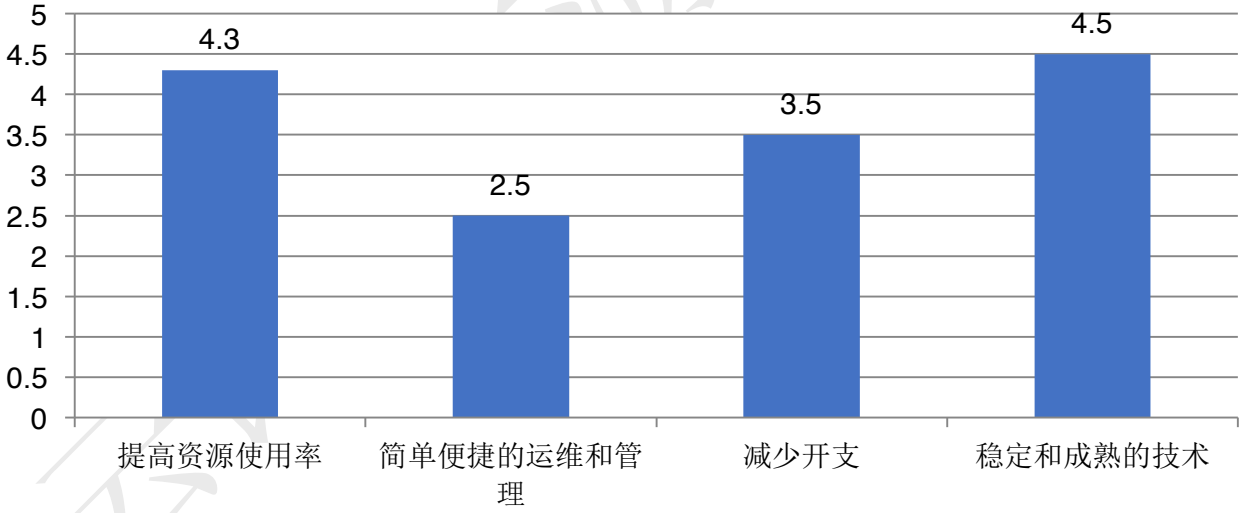

图12 用户选用 Meoso、DC/OS 的主要原因（资料来源：2017 年中国 Mesos、DC/OS 调研报告）

使用 DC/OS 可以为企业收入带来持续增长，其中，移动互联领域增长速度最快，预计到 2017 年第二季度，总收入将超过 12 亿元人民币。此外，电信、金融等行业也将得到稳步增长。

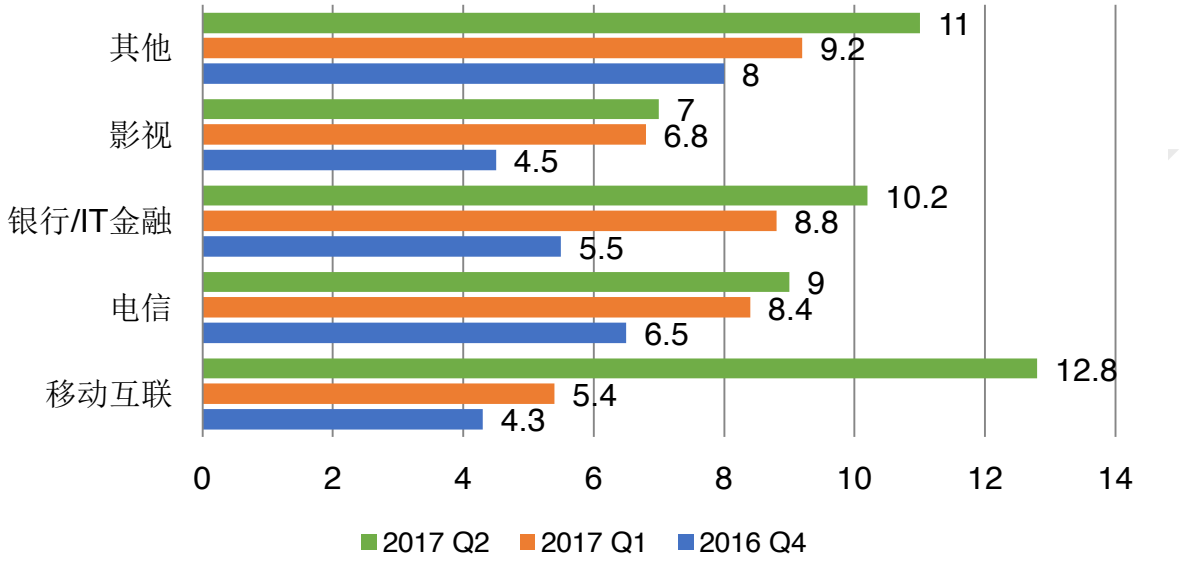

图13 DC/OS 用户收入增长（资料来源：2017 年中国 Mesos、DC/OS 调研报告）

（二） DC/OS 中国产业发展面临的挑战

根据《2017 年中国 Mesos、DC/OS 调研报告》发布的结果，用户当前对于使用 DC/OS 困惑主要在于人才的稀缺、安全机制以及缺少相关工具。

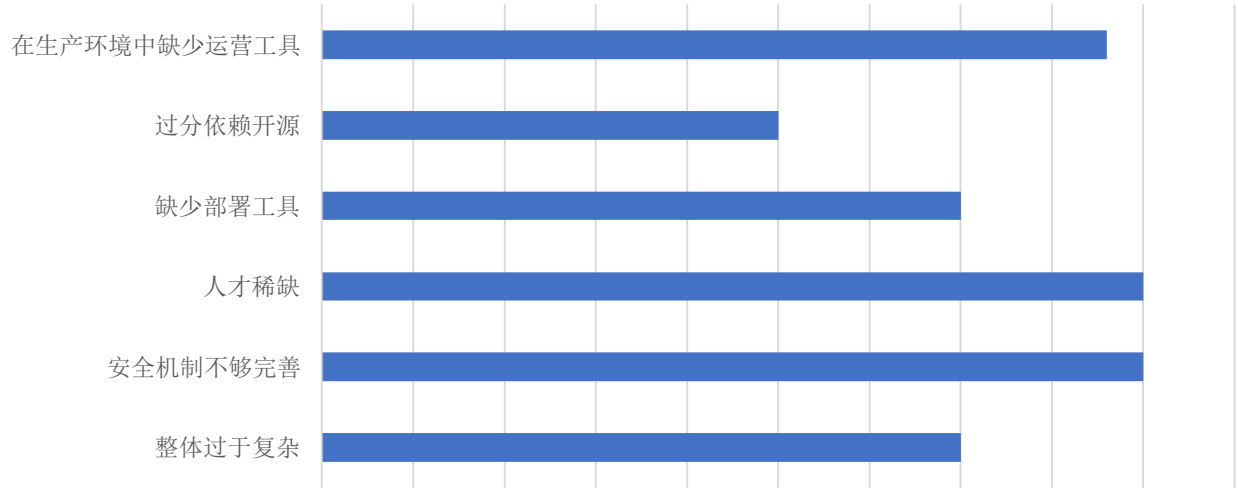

图14 用户使用 DC/OS 的主要困惑（资料来源：2017 年中国 Mesos、DC/OS 调研报告）

在技术层面

在调研结果中，缺少运营和部署工具、安全机制和技术复杂方面的问题是用户的主要顾虑。

在工具方面，DC/OS 官方提供的工具较少，主要是 Bootstrap、Executor、Scheduler 等，用于安装、执行和调度。同时，DC/OS 还支持全部基于 Mesos 框架开发的工具，开发者可以自行开发所需的工具。目前，DC/OS 已经制定了发展路线（Roadmap），确定将会推出官方 SDK，以便更容易得编写新服务或者集成现有服务。

在安全方面，一方面安全技术需要不断完善，虽然 DC/OS 已经提供了基于 Linux 的隔离策略，选用 Marathon 作为容器调度引擎，控制 Cgroups、Docker Container 中的微服务。与容器安全机制类似，当 DC/OS 直接运行在裸机之上时，不可信的服务可以直接对物理资源造成威胁，并对整个 DC/OS 系统稳定造成影响。另一方面，市场上普遍存在对开源技术的偏见，认为基于开源技术的产品普遍存在大量安全隐患，攻击者可能从公开的源码中获取攻击或注入点，对产品安全稳定造成威胁。

在市场推广层面

根据《2017 年中国 Mesos、DC/OS 调研报告》，中国市场对 DC/OS 的了解程度较低，有超过半数的受访者表示从未听说过 DC/OS 技术；仅有 1%的受访者表示已经将 DC/OS 投入到生产环境中。

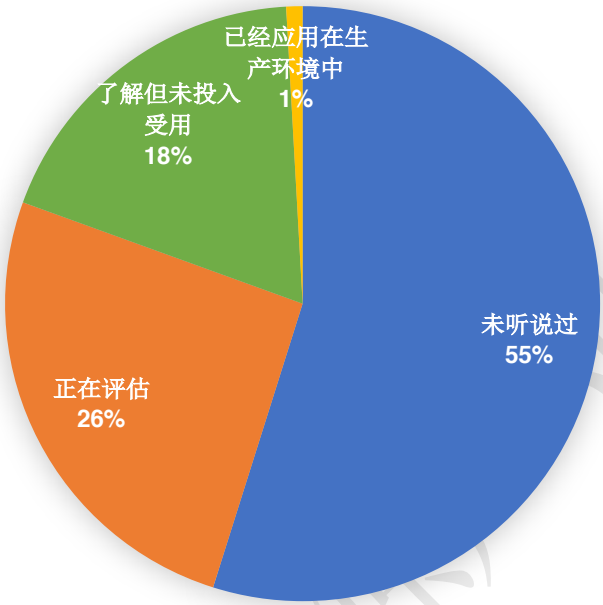

图15　中国市场对 DC/OS 的了解程度（资料来源：2017 年中国 Mesos、DC/OS 调研报告）

对此，DC/OS 社区和云计算开源产业联盟已经制定了下一步工作计划，即加强在中国的技术推广工作，提高 DC/OS 的市场影响力，生成更多成熟可靠的产品应用案例。

在标准制定和人才培养方面

由于 DC/OS 进入市场的时间较短，所以缺少相应的人才培养和积累。目前，市场上也没有完善的 DC/OS 相关培训体系，人才培养主要依靠社区组织的技术沙龙和研讨活动。此外，相关的产品技术标准和应用场景规范尚未完全建立，后续还需要继续完善，推动市场规范和产品标准化。

六、 DC/OS 发展趋势预测

（一）将会持续推出成熟、优化的技术产品

➢ 短期发展——快速提升技术易用性和兼容性

Pods

Pod 是一个轻量级的节点，同一个 Pod 中的容器可以共享同一块存储空间和同一个网络地址空间。Pods 支持“sidecar pattern”（挎斗模式）容器技术。挎斗模式第一种单节点多容器模式，主要利用在同一 Pod 中的容器可以共享存储空间的能力。

标准的 API

DC/OS 将发布自己的 API 标准，用于支持自身整合工具集，如 Graphite、 Grafana、 InfluxDB 和 Prometheus 等。

整合的日志系统

包括 Master、Agent 和服务在内的 DC/OS 所有组成部分的日志都将通过 Journald 进行汇总，并提供给日志整合系统进行分析，如 Splunk、ELK 等。Journald 是 Linux 新系统日志方式，采用这种方式的日志写入二进制文件中，可以按照管理员的需求输出指定格式、样式的日志。

支持 GPU 的发现、隔离和功耗

DC/OS 提供的隔离方法保证不同的容器之间不能进行交互，依赖此种技术，用户可以为不同的任务请求独立的 GPU 支持。

兼容 Windows 系统

DC/OS 将在短期内支持 Windows 操作系统，并支持通过 Windows 容器调度 DC/OS 资源。

➢ 长期规划——对大规模集群管理的不断优化

调试工具

不断加入用于运行时的调试工具，并与现有调试工具（如 gdb 和 IDE）集成。

自动伸缩

根据既定的策略（如响应时间和吞吐量），自动横向扩展基于容器的应用程序。

（二）稳步提升市场普及率

提高市场普及率是推动 DC/OS 技术落地产业的最直接手段。目前，DC/OS 在中国市场的普及率远低于 OpenStack、Docker 等技术，虽然已经在 Twitter、Facebook 等国外大型互联网企业落地，但在国内只有少数的成功案例，了解 DC/OS 的企业多数也处在观望和评估阶段。对此，DC/OS 社区和 Mesosphere 已经有针对性的提出了中华区市场推广策略，其中就包括从 2016 年开始的发布会和 Mesos con 大会，下步还会继续在中国大陆的其他地区开展相关的推广和发布活动。

（三）统一标准和应用场景的制定

制定统一的产品标准是规范产品和市场的重要手段，对此，云计算开源产业联盟已经于 2016 年发布了可信云·开源解决方案评估，推出了面向 OpenStack 解决方案的评估标准和方法。目前，联盟已经计划制定面向 DC/OS 技术的评估方法，以此方式给用户采购 DC/OS 解决方案提供参考，也给厂商开发产品提供规范。

同时，为了解决 DC/OS 在产业落地中遇到的客户需求和技术场景困惑的问题，云计算开源产业联盟于 2016 年发布了开源项目，其中就包括 DC/OS 应用场景这一项目，用于研究在云计算的实际部署中，不同的行业表现出的不同需求，如金融行业用户注重风险可控、两地三中心，广电行业用户侧重大量的流媒体处理能力等。随着 DC/OS 在

产业中的不断落地，联盟也将不断完善此项目，形成完整的标准文稿，供产业界参考。

七、 DC/OS 应用案例

中国联通基于 DC/OS 的多租户容器化调度管理平台方案

针对 DC/OS 的特点和联通的业务特性，联通提出了基于 DC/OS 的多租户容器化调度管理平台方案。如图所示。

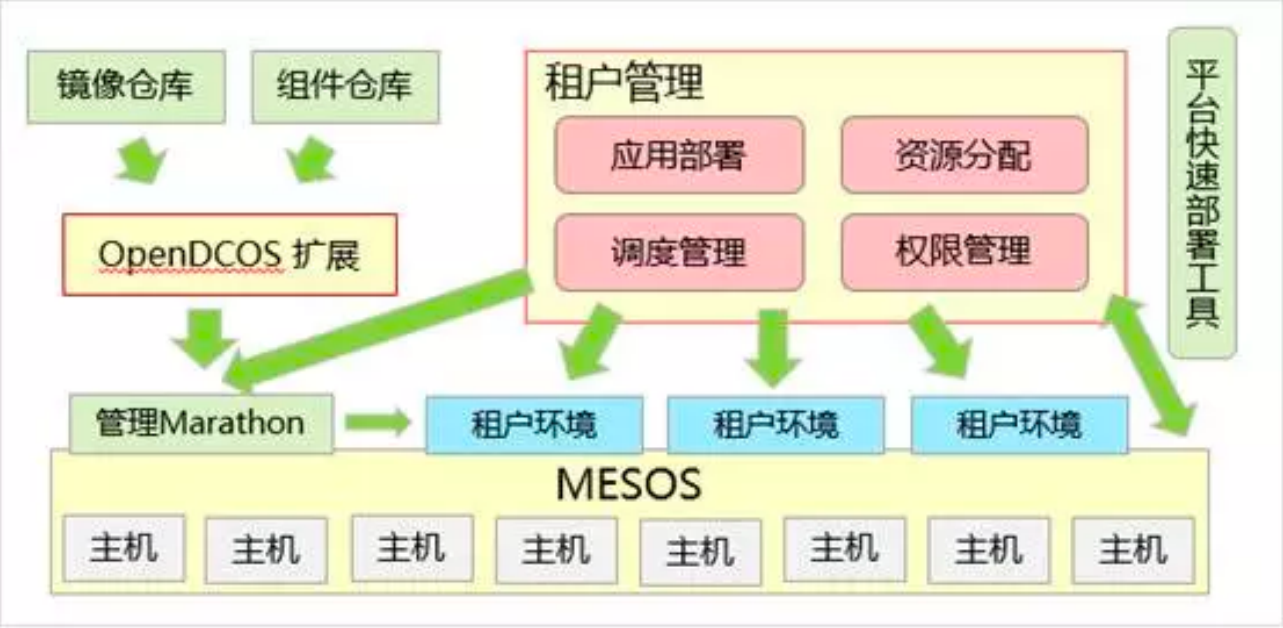

图16 中国联通 DC/OS 容器化平台方案架构

具体的改造点如下：

租户资源分配

以 Mesos 为核心，通过改造 Mesos，进行安全加强，利用 Mesos 的资源分配，提供灵活的租户资源分配和回收的基础能力，并通过改造 Admin Router 对外提供 REST API 接口。

租户应用管理调度

通过每个租户分配独立的 Marathon，提供租户下的应用生命周期管理。各租户之间的应用由不同的 Marathon 独立管理调度。

租户应用部署

租户应用部署仍然支持原始 Open DC/OS 的两种方式：直接部署镜像仓库里的容器和从组件仓库部署组件包。由于当前 OpenDC/OS 版本无租户概念，内置组件仓库模块需要改造，以适应多租户环境。

权限管理

通过改造 Mesos 及 Open DC/OS，并利用 OpenDC/OS 的 OAuth 能力，贯通整个多租户管理平台的用户认证和授权。

效果：

通过如上改造，OpenDC/OS 的能力有了极大的提高，已经可以适应软研院对容器化平台的需求，通过该平台可以打造新一代的软件生态环境。利用平台的资源分配能力、应用管理能力、通用组件能力，构建软件全生命周期支撑平台。

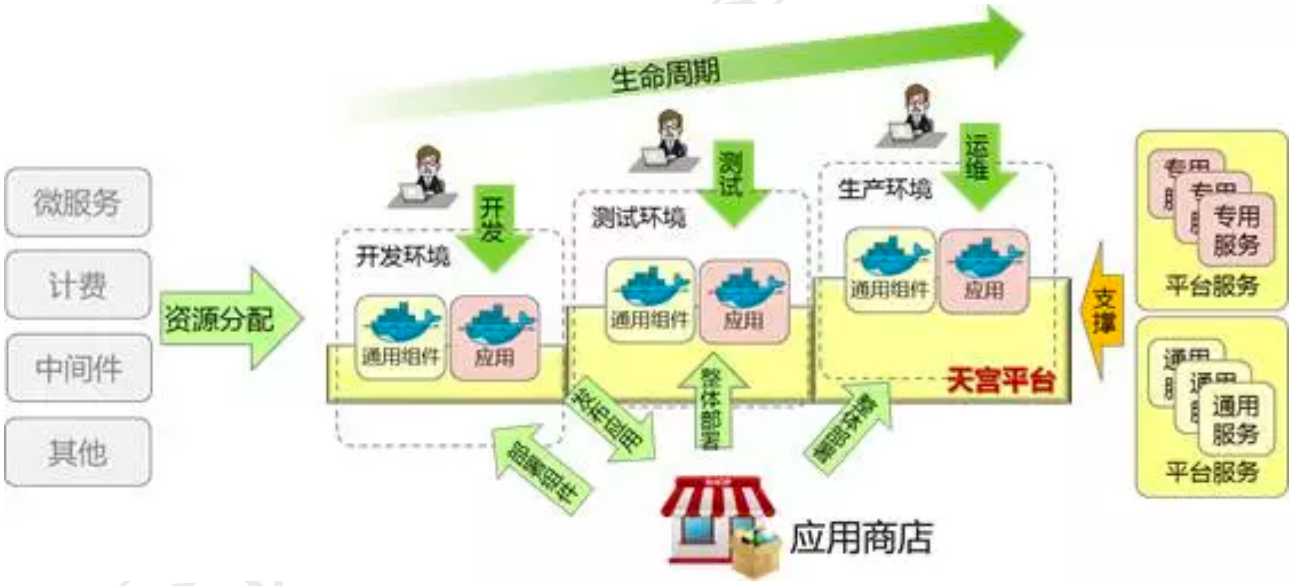

图17 中国联通 DC/OS 容器化平台方案应用示意

使用该平台可以做到：

- 大规模集群主机动态管理
- 多租户细粒度的动态的资源调度分配，灵活的资源控制策略；

- 为应用提供隔离的、安全的运行环境，并在一点提供可视化终端、日志查询等应用管控机制
- 应用秒级安装部署，支持静默安装、可视化配置、自定义配置等多种部署方案；
- 支持服务实例的实时弹性扩展，动态部署；

目前该平台已经在软研院内部的若干项目中使用（cBSS、eSIM、总经理在线等），取得很好的效果。经过统计使用该平台后，提高整体资源利用率 10 个百分点，节约大量运维工作量、人工成本，节省 cBSS 等项目主机投资 1500 万。

数据中心操作系统作为新的 IT 基础设施，对于互联网相关企业的重要性不言而喻，建设一个支持快速部署服务，同时保持敏捷、高效、安全和服务质量的现代数据中心平台，对于电信运营商既是需求也是挑战。DC/OS 提供了一个生产可行的方案，运营商根据业务特点，选择合适的业务分步进行改造和迁移，最终对数据中心乃至全网 IT 系统进行 DC/OS 改造，是一个相对安全可靠的选择。通过推广该平台可以为各系统节约大量的硬件投资及开发运维成本。

云计算产业联盟

烽火通信楚天云平台案例

楚天云在 2016 年 6 月正是上线之后，客户又提出了新的 PAAS 需求，烽火通信为此专门建立了单独的网络区域来部署 DC/OS 平台，其部署架构见下图：

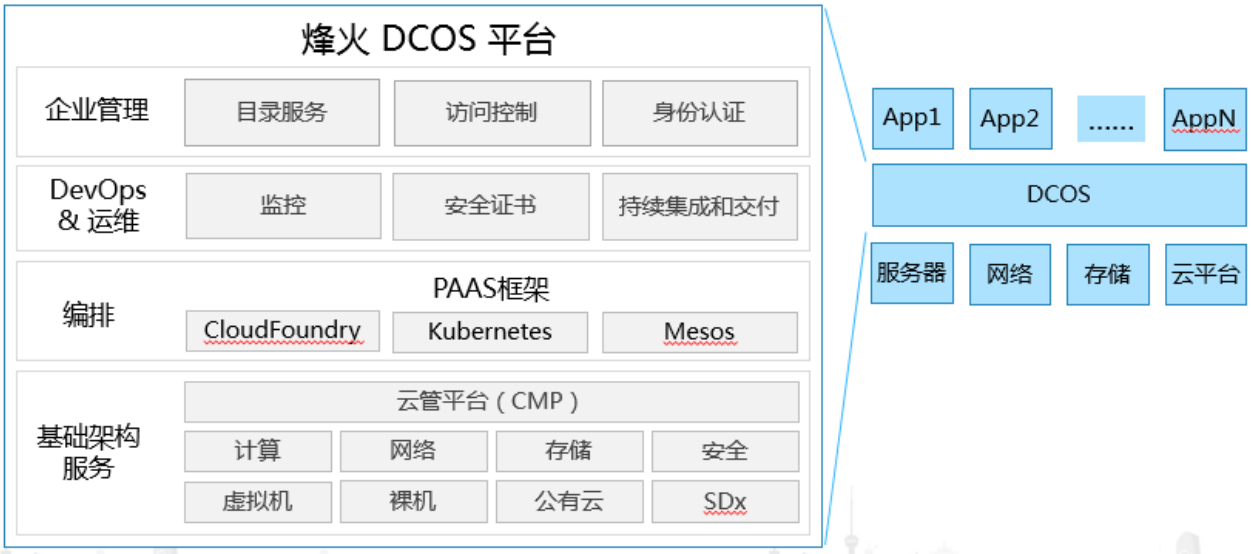

图18 烽火楚天云 PAAS 平台 DC/OS 部署架构

通过部署烽火楚天云 DCOS 平台，可以对应用提供如下服务：

a) 一键部署应用

可以自定义虚拟应用和容器应用的参数，通过图形化界面实现一键式自动化部署，并且提供快速回滚到历史版本；

b) 服务发现

新的实例自动加入到现有应用集群人工干预，实现应用无缝扩展；

c) 弹性扩缩

以云资源 CPU、内存为触发阀值，自动进行应用实例扩缩和负载均衡，应对业务高并发需求，保证用户体验；

d) 持续集成

可连接公共或私有代码库，通过自动化的方式进行代码构建，包括编译、打包、测试和发布，提升软件开发及部署效率；

e）灰度发布

调整不同应用镜像的灰度发布比例，逐步过渡用户访问到新版本应用，大大降低由于新版本问题导致对整个业务的影响。

另外烽火楚天云 DCOS 平台同时对接下层虚拟化平台，实现资源调度：

a）分布式资源调度

为大数据集群、数据库集群、容器集群按需调度资源，支撑集群的按需扩展；

b）动态任务调度

采用任务调度器实现自动化容器调度、应用管理、动态部署应用和服务，并提供应用运行监控；

c）弹性扩展

提供物理集群、虚拟注意、虚拟硬盘、网络等资源的在线弹性调整；

d）在线迁移

采用高性能分布式存储，实现虚拟主机的安全迁移，保证业务不中断；

e）跨数据中心管理

支持不同数据中心的云资源统一管理，可根据应用需求进行分区域运行。

云计算开源产业联盟