

OPEN SOURCE AND FREEDOM

龙辉 2018.10

微博、微信：@huihoo

<https://huihoo.com>

议题

一: Open. For Business.

二: 算法、算力、数据

三: 用户端(B端、C端)思考

四: 编程语言、操作系统、数据库

五: 投资组合

六: 开源治理、企业研发

OPEN SOURCE

Freedom

自由软件、自由社会

Open. For Business.

开源已成为企业的核心竞争力

Google, Facebook, Microsoft等巨头们正在GitHub上展开开源的军备竞赛。

建立开源生态系统

巨头们都在开源，力求在某个领域建立领导地位

如Google, Chrome和Android取得巨大成功，现在的TensorFlow和Kubernetes正在续写成功。

选择开源软件

选择开源实质是选择生态和站队，当然解决问题是首要考虑

之前，有篇文章 [《值得关注的开源软件》](#) 大家可参考

值得关注的开源软件推荐

给出一些思路：系统化的了解、学习、搭建软件堆栈和开放平台

- 一、关于这篇文章
- 二、开源软件很多很多
- 三、第一类：框架和编程驱动的应用层
- 四、第二类：数据驱动的大数据、人工智能层
- 五、第三类：云计算、基础设施层
- 六、第四类：操作系统、编译器、虚拟机、编程语言
- 七、45个开源软件特性、要点
- 八、资源汇总

开源硬件

之前，我们多谈软件的开源，这次我想更多聊聊硬件和汽车工业开源的事，当然这些项目的核心还是软件。

从芯片、服务器、存储、机器人到自动驾驶都有开源生态系统，我们现在都能通过开源项目去接触和了解这些核心技术。

算法、算力、数据

人工智能

芯片、大数据

计算工业说简单点，就是这三点

而在这三方面拥有核心能力和领导力的都将是行业巨头，如：GOOGLE

它拥有强大的研发和算法能力，拥有庞大的数据中心和TPU，拥有海量的数据和存储。

算法：开源机器学习、深度学习提供算法基础

算力：Open Compute Project + RISC-V 提供了算力可能

数据：Apache数据流、大数据生态系统

Open Compute Project

FACEBOOK 要通过此项目重新定义数据中心，并挑战 GOOGLE、AMAZON 在数据中心的领导地位。

Microsoft也是该计划的积极参与者

Microsoft Project Olympus

v1.0 Contribution OCP/DCD Nov 8, 2017

PROJECT OLYMPUS – OPEN SOURCE HARDWARE

Next-gen Cloud Hardware

Open sourcing leading edge Hyperscale cloud hardware currently under development at Microsoft

Development Model

New collaboration model with OCP community – co-develop open hardware at cloud speed

Industry Ecosystem

Bootstrap a vibrant ecosystem in OCP for the next generation of datacenter hardware

<http://www.opencompute.org/wiki/Server/ProjectOlympus>

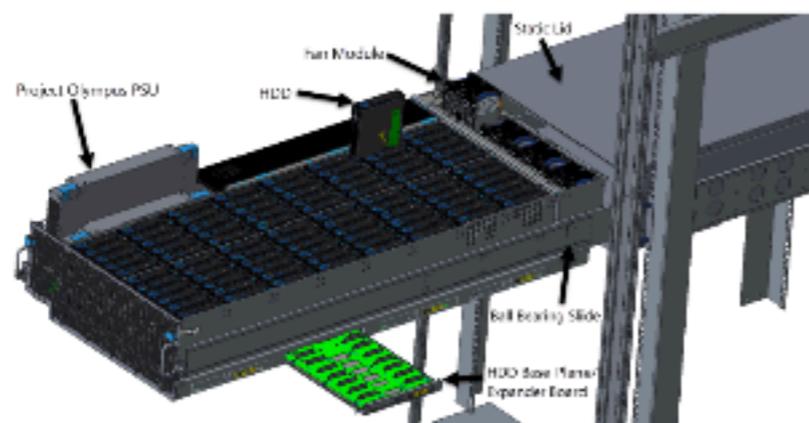


Microsoft Project Olympus

Expansion Chassis Spec Contributions

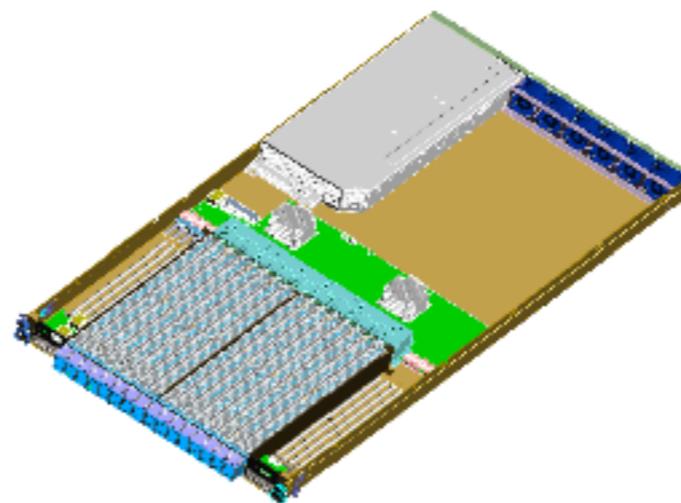
DX-88 DISK EXPANSION

88 Hot-Swap Hard Disk Drives



FX-16 FLASH EXPANSION

16 Flash Modules
256+TB



HGX-1 GPU EXPANSION

8 NVLINK ENABLED GPUS



RISC-V

自由、开放的 **RISC** 精简指令集架构，采用**BSD**许可证

有望成为ISA(CPU)领域的**LINUX**

该项目2010年始于加州大学柏克莱分校，David Patterson教授带领团队开始从头开始设计一个干净的ISA。经历了很多年，经过多次流片验证，终于在2014年发布了最终版spec，就是RISC-V（V是第五代的意思）。

《计算机体系结构：量化研究方法》

JOHN HENNESSY 和 DAVID PATTERSON 在计算机架构领域可谓是无人不晓，他们的成就包括 MIPS, RISC, RAID, 以及两人合著的这本书



JOHN L HENNESSY 

United States – 2017

CITATION

For pioneering a systematic, quantitative approach to the design and evaluation of computer architectures with enduring impact on the microprocessor industry.



RESEARCH

前斯坦福大学校长 John L. Hennessy



DAVID PATTERSON 

United States – 2017

CITATION

For pioneering a systematic, quantitative approach to the design and evaluation of computer architectures with enduring impact on the microprocessor industry.



RESEARCH

加州大学伯克利分校退休教授 David A. Patterson

Google 的母公司 Alphabet 已宣布 John Hennessy 接替 Eric Schimdt 就任董事长。而在这之前，David Patterson 就已经开始在 Google 担任杰出工程师，主要从事硬件架构方面的研究工作，Google 的 TPU 2 就有 Patterson 的巨大贡献。

两位宗师再聚首，会师 Google，人工智能时代体系结构创新令人期待

Domain Specific Architecture

DOMAIN SPECIFIC ARCHITECTURES (DSAs)

- Achieve higher efficiency by tailoring the architecture to characteristics of the domain
 - Not one application, but a domain of applications
 - Requires more domain-specific knowledge than general purpose processors need
- Examples:
 - Neural network processors for machine learning
 - GPUs for graphics, virtual reality
- Some good news: demand for higher performance focused on such domains

2017年3月，Hennessy在斯坦福发表演讲，提出了目前处理器的新希望在于 Domain Specific Architecture (DSA，即针对应用领域做优化的处理器架构，区别于通用架构)。Hennessy在演讲中说，DSA的优点是可以为特定的一类应用做架构优化，从而实现更好的能效比。相对于通用处理器，DSA需要在设计时考虑专用领域的特殊需求，也需要设计者能对该领域有深入的理解。而DSA的一个绝佳例子，就是为机器学习设计的神经网络处理器。DSA设计将会成为处理器架构的新趋势。

谷歌的 DSA —— TPU

思考: Domain Specific Language (DSL) 领域特定语言

用户端(B端、C端)思考

企业应用

终端消费

老客户、订阅服务、会员制

不管做什么，首先要活下来

编程语言、操作系统、数据库

回归本质和源头

认清事物本质，洞察未来

编程语言

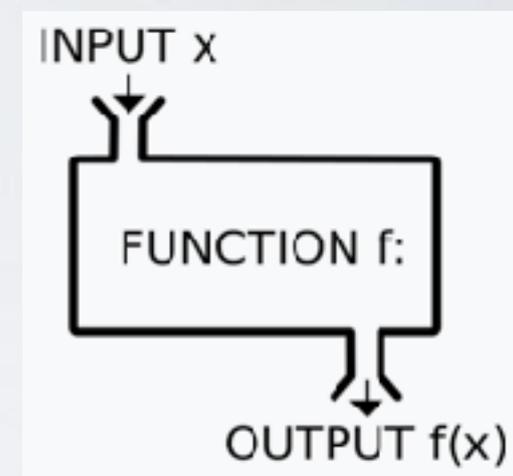
面向对象编程语言、函数式编程语言

对象是一切，数学函数

克莱因认为函数为数学的“灵魂”

函数 $f()$ 就像机器或黑盒，给予输入值 x 便产生唯一输出值 $f(x)$ 。

Erlang、Scala、Clojure、Haskell ...



操作系统

可在移动、嵌入式、实时操作系统寻找机会和更多沉淀

个人倾向移动操作系统，简单讲就是在Android的分发版下功夫，因为手机是距离我们隐私最近的设备。

LineageOS、Replicant两个Android分发版可关注。

AGL、ROS：机器人、汽车自动驾驶

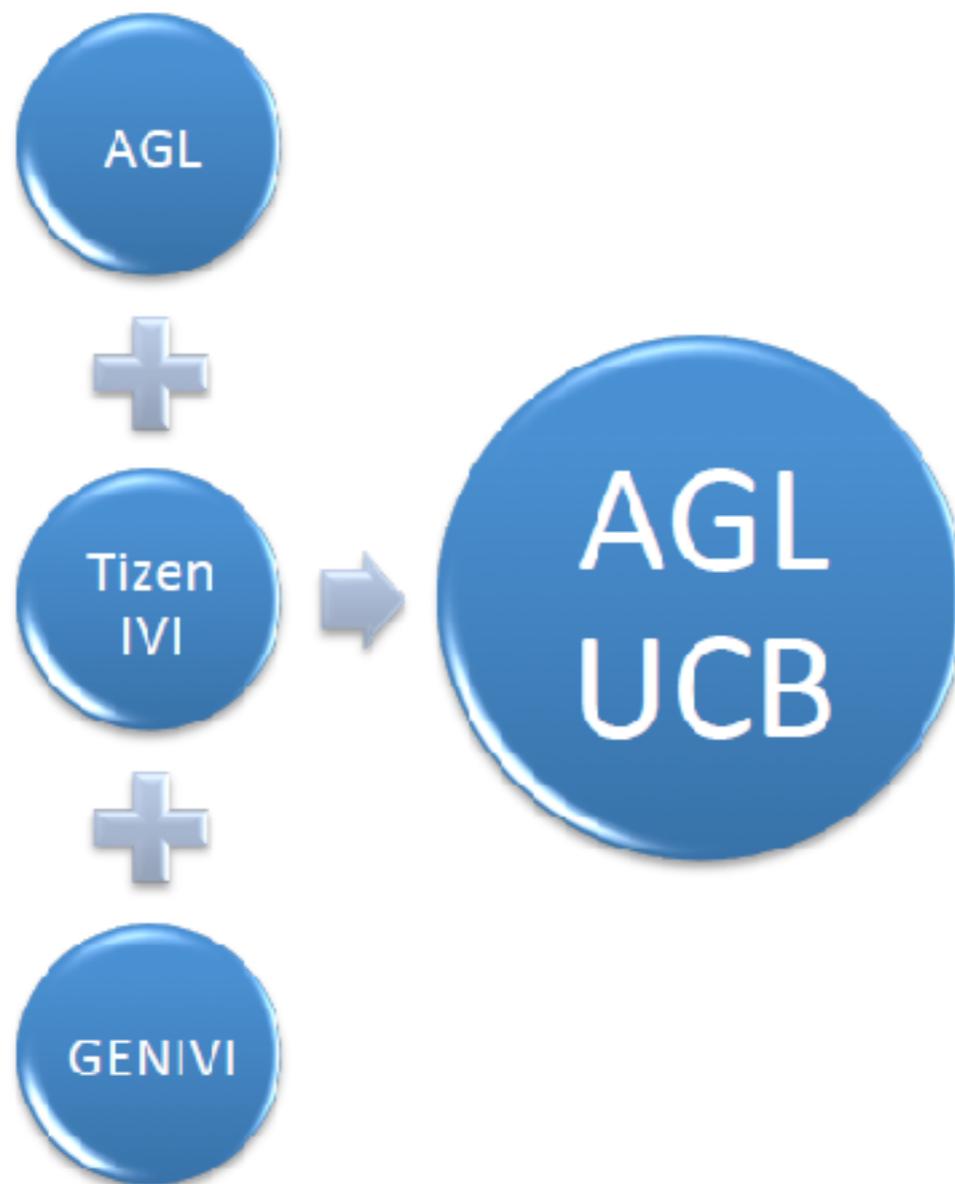
Autoware.ai 和 Comma.ai 也膜拜下天才

AUTOMOTIVE GRADE LINUX(AGL)

AGL基于Tizen项目构建，最初由韩国三星主导，现在为Linux基金会项目。

基于AGL的丰田凯美瑞(Toyota Camry)将于2018亮相北美市场

AGL Distro “Unified Code Base”



- *First Release announced at CES Las Vegas in January*
- *Unifying the best of AGL, Tizen IVI and GENIVI into a single code base for the entire industry!*
- *Reduce fragmentation, focus on innovation and new features!*
- *Yocto/Poky based with AGL specific layers*

2018 Toyota Camry with Entune 3.0 and AGL



ROS

ROS机器人操作系统，这几年在自动驾驶领域也有了长足发展。

百度Apollo(阿波罗)就是基于ROS构建，大家可更多关注另外一个基于ROS的自动驾驶开源项目。

Autoware.ai

Autoware

Sensing

Camera

LiDAR

IMU

GPS



Computing

Perception

Localization

Detection

Prediction

Planning

Mission

Motion

Decision

Intelligence

State

Actuation

YMC

AS

ZMP

Custom



Data

Socket

System

Util

自动驾驶核心算法：SLAM

SLAM 全称 Simultaneous Localization and Mapping，同时定位和地图构建

如果你不懂SLAM，那最好还是不要说你懂自动驾驶汽车了。

Computer Vision

COMMA.AI

不到1000美金的自动驾驶后装套件

Comma.ai 自己研发的基于深度学习的自动驾驶软件，是一个“对于特斯拉的 Autopilot 辅助驾驶技术的替代方案”

<https://github.com/commaai/openpilot>

自动驾驶后装套件



EON Gold Dashcam DevKit

Retrofit your car with a copilot.

\$499.00



Added Panda OBD-II Interface (+ \$199.00)

Grey (High Precision GPS)



Added Giraffe Connector (+ \$60.00)

Toyota



支持本田(Honda)和丰田(Toyota)

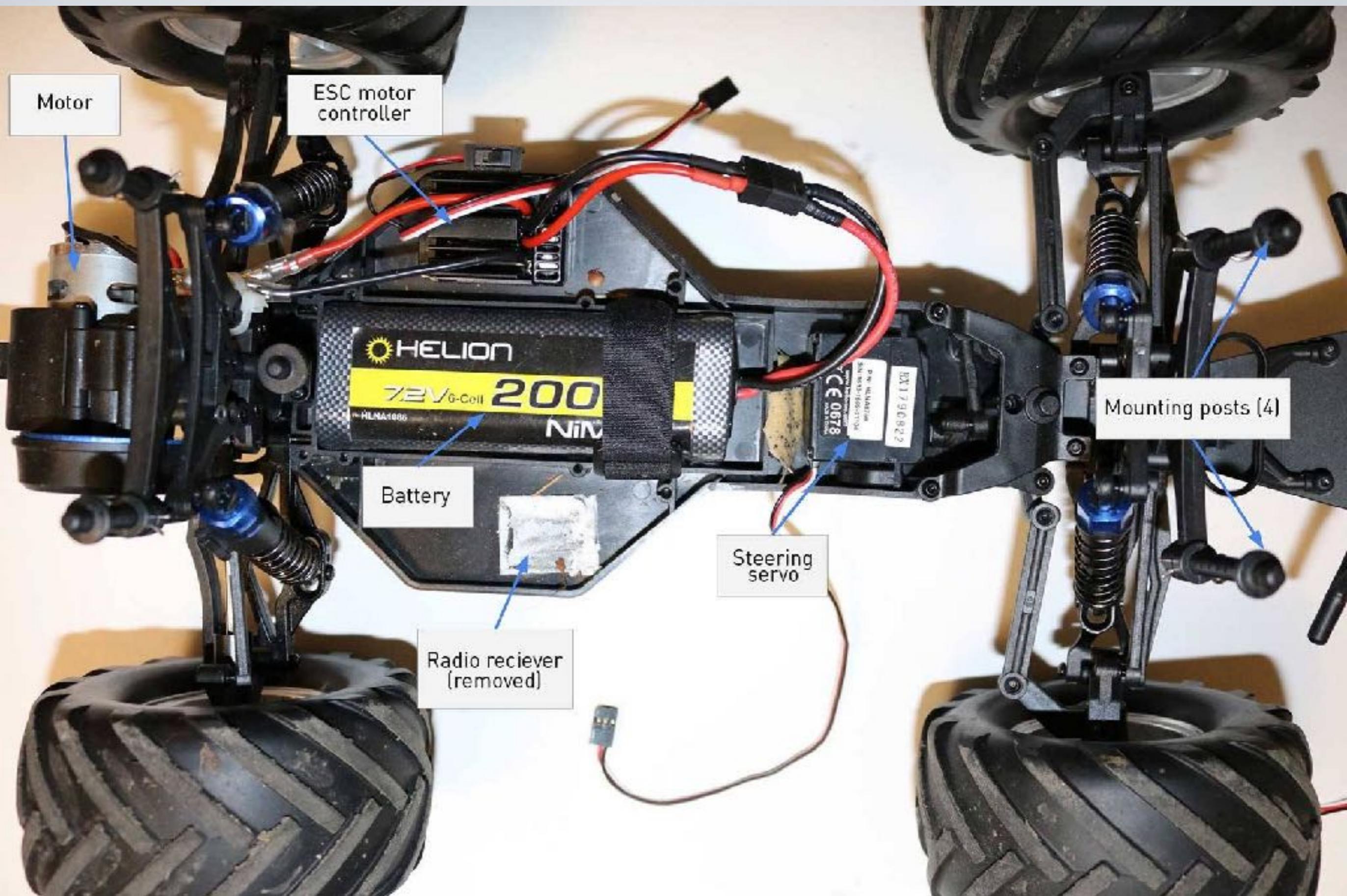
DONKEY CAR

刚才介绍的这些开源自动驾驶项目，相信我们大多数人都没有机会去接触这些酷酷的技术，但又想学习，怎么办？

Donkey Car开源项目给我们提供了一个学习相关技术的入门途径，它有趣、低成本，可以带着孩子一起探索、一起玩。

<http://donkeycar.com/>





Motor

ESC motor controller

HELION
7.2V 6-coil 200
NiMH
HLMA1006

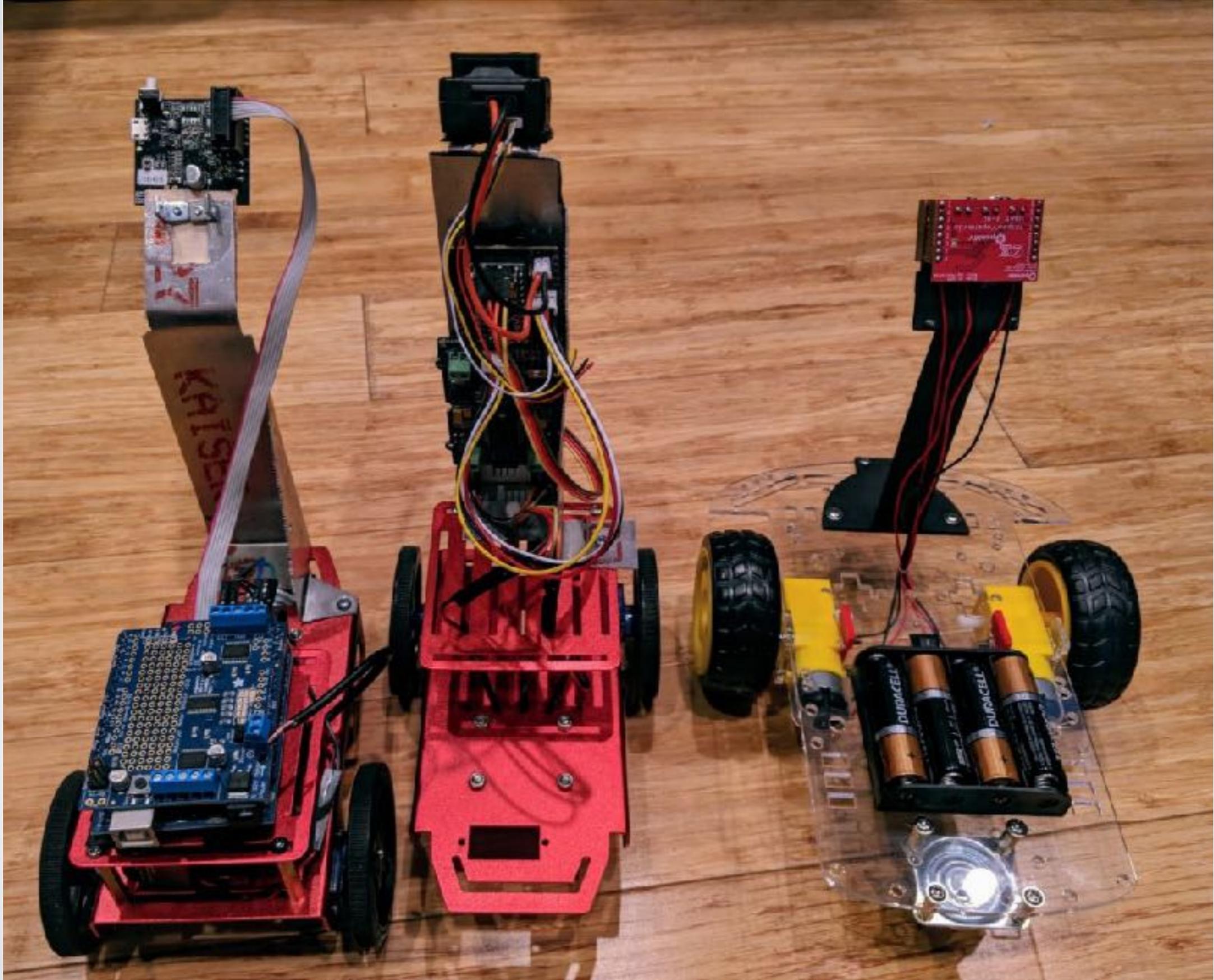
Battery

Radio reciever
(removed)

Steering servo

Mounting posts (4)

Donkey Car小车是基于树莓派和Python构建的，利用Keras深度学习框架训练和实现自动驾驶小车，是一个Python驱动的玩具车。Donkey Car提供了完整的软件、硬件清单，你可购买套件并完成硬件组装和软件安装。



Donkey Car能抓取、分析训练数据集， 这点很牛逼

Donkey



0

1x playback

▶ Play (—)

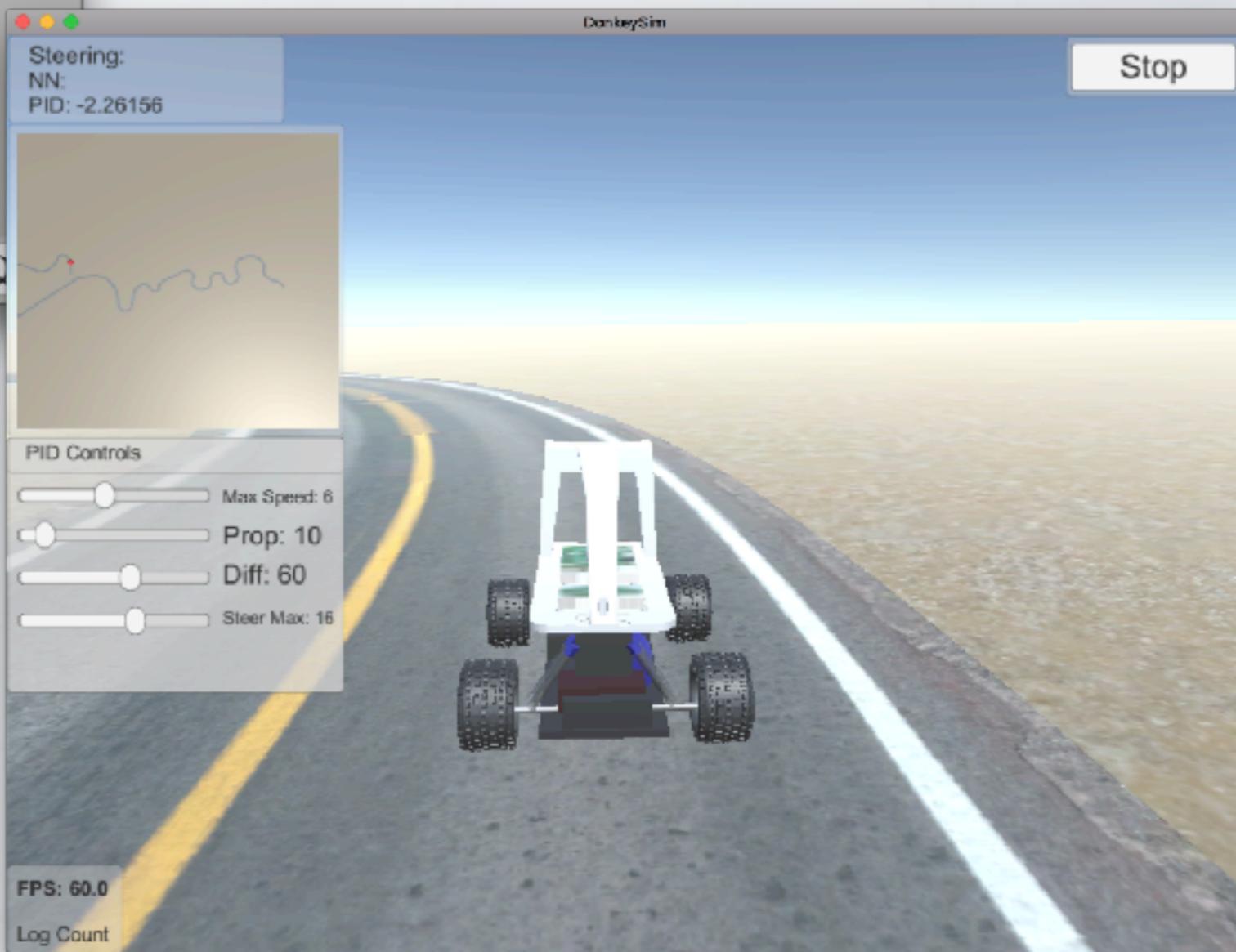
◀ Rewind 1s (b)

✂ Split (c)



Save and Delete Checked Training Data!

Donkey Car还提供了模拟器软件，你可以直接下载去感受下



数据库

可建立以Hadoop和RDBMS组合的数据存储基础设施

SQL on Hadoop, Spark on Hadoop

基于HADOOP的OLTP/OLAP、SQL引擎、数据库存储和数据仓库解决方案

Apache Spark、Spark on HBase & Spark on YARN

Apache HBase、Apache Phoenix

Apache Hive、Apache Kylin

.....

一个参考架构：

Elasticsearch和Hortonworks作为数据的核心支撑，Metabase是BI和分析门户。

投资组合

人生苦短，我用Python :) 我们需要平衡好自己和公司的技术投资

个人喜好，可参考：

职场：4C + Go (容器云/DevOps)

秘密武器：Erlang(原子语言) + Julia(科学语言)

Rust：因为火狐，❤️🦊。

4C语言：C、C++、C#、Clojure，考虑两大虚拟机平台：.NET和JVM，一个平台选一个。

ERLANG

Erlang在笔记本上轻松创建了300万个进程

```
deepbrain:code huihoo$ erl +P 3000000
Erlang/OTP 19 [erts-8.3] [source] [64-bit] [smp:4:4] [async-threads:10] [hipe] [kernel-poll:false] [dtrace]

Eshell V8.3 (abort with ^G)
1> processes:max(500000).
Maximum allowed processes:4194304
Process spawn time=3.4 (7.772) microseconds
ok
2> processes:max(1000000).
Maximum allowed processes:4194304
Process spawn time=3.61 (7.843) microseconds
ok
3> processes:max(2000000).
Maximum allowed processes:4194304
Process spawn time=3.985 (8.2935) microseconds
ok
4> processes:max(3000000).
Maximum allowed processes:4194304
Process spawn time=4.1333333333333334 (8.520333333333333) microseconds
ok
```

JULIA

Julia是一个面向**科学计算**的高性能动态高级程序设计语言，采用MIT许可证。除了语言本身的优点，**Julia**还拥有非常强大的生态系统，主要应用于数据可视化、通用计算、数据科学、机器学习、科学领域、并行计算六大领域。

application	Julia	Python	R
fib	0.91	30.37	411.36
quicksort	1.14	31.89	524.29
mandel	0.85	14.19	106.97
pi_sum	1.0	16.33	15.42

个人选择Julia更多是因为数学、物理，作为两大基础科学强大的语言工具。

RUST

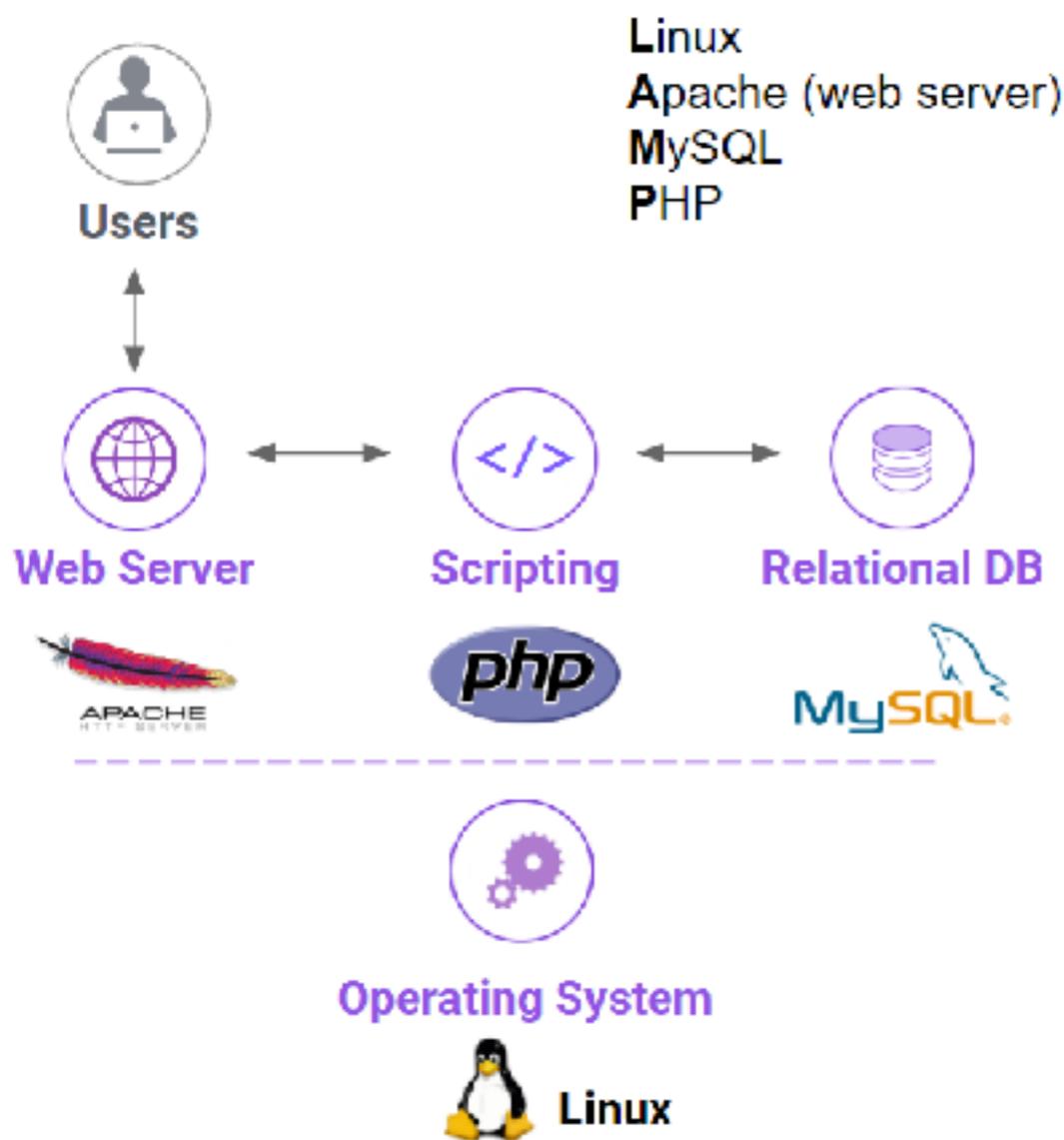
一种更安全的编程语言，火狐新版Firefox Quantum已内置采用Rust编写的新引擎。



Rust Comes to RISC-V

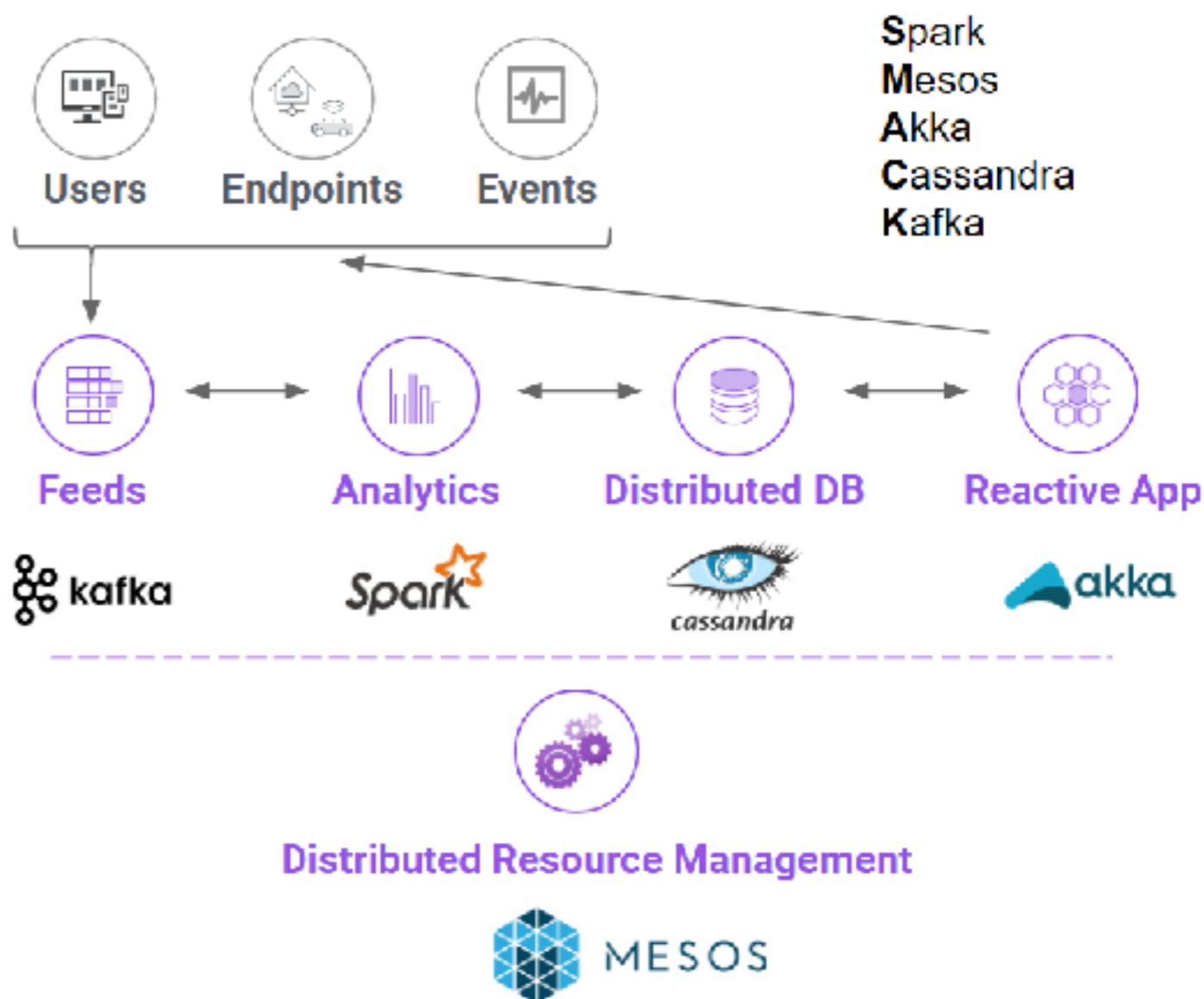
LAMP Stack

Enabling dynamic web applications



SMACK Stack

Powering scalable real-time & data-driven applications



从LAMP堆栈到SMACK堆栈的演进



MESOS



简单、清晰的架构层次

Engine: Apache Spark

Manager: Apache Mesos

Model: Scala and Akka

Storage: Apache Cassandra

Broker: Apache Kafka

计算基础设施

为端对端业务提供实时计算服务

新一代大数据解决方案

内存计算、一个思考方式、一种编程模型

Apache Spark

流处理、交易系统、广告系统、数据分析、机器学习、多语言接口

高并发、分布式、容错、事件驱动基础设施

参与者模型(ACTOR MODEL)、一种编程模型

参与者模型推崇的哲学是“一切皆是参与者”，这与面向对象编程的“一切皆是对象”类似，但是面向对象编程通常是顺序执行的，而参与者模型是并行执行的。

Akka

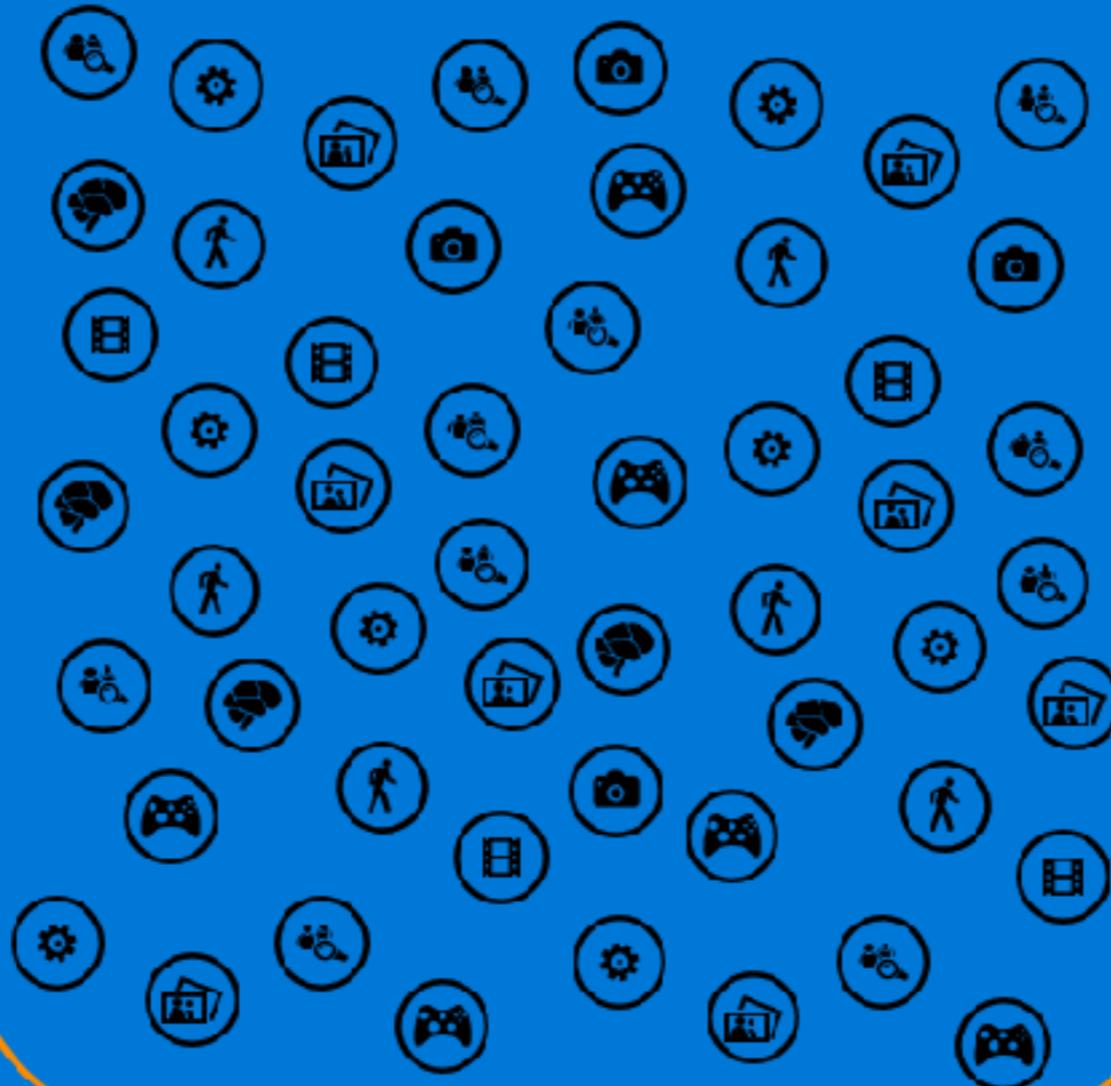
Actor模型的本质就是消息传输。作为一种计算实体，Actor与原子类似。

ACTOR的原子世界

Frontends



Actor-based Middle Tier



Storage



分布式流式处理基础设施

连接端对端设备应用和基础设施的HUB

快、解耦、高可用

生产消费模型、一个思考方式、一种编程模型

Apache Kafka

流处理、业务系统、交易系统、多行业、BaaS、多语言接口

分布式存储基础设施

为业务运营提供存储服务

分布式存储服务，位于存储系统层

Apache Cassandra

时序数据、用户数据、设备数据、权益数据、日志消息、多语言接口

分布式集群基础设施

提供分布式集群服务

SPARK、AKKA、CASSANDRA、KAFKA、DOCKER集群管理

Apache Mesos

数据中心、可扩展、容器云、分析云、云数据库、云消息

DC/OS仪表盘

← → ↻ m1.dcos/#/catalog/packages?_k=b89g6c ☆ [extension icons]

DC/OS ihuihoc@gmail.com

- Dashboard
- Services
- Jobs
- Catalog**

RESOURCES

- Nodes
- Networking
Networks

SYSTEM

- Overview
- Components
- Settings
- Organization

Catalog

 cassandra CERTIFIED	 chronos CERTIFIED	 confluent-kafka CERTIFIED	 datastax-dse CERTIFIED
 datastax-ops CERTIFIED	 dcos-enterprise-cli CERTIFIED	 elastic CERTIFIED	 hdfs CERTIFIED
 jenkins	 kafka	 kibana	 marathon

K8S ON DC/OS

Services >  kubernetes Running (1) 

Instances Configuration Debug

<input type="checkbox"/>	ID	NAME	HOST	STATUS	HEALTH		CPU	MEM
	kube-controller-manager-2-instance_ec1f200b-19...	kube-control...	192.168.0.240	Running	●		0.5	512 MiB
	kube-controller-manager-1-instance_6150bb1f-a9...	kube-control...	192.168.0.241	Running	●		0.5	512 MiB
	kube-controller-manager-0-instance_41d1807b-7...	kube-control...	192.168.0.242	Running	●		0.5	512 MiB
	kube-apiserver-0-instance_e658ba1f-fedf-4d84-...	kube-apiserv...	192.168.0.241	Running	●		0.5	1 GiB
	kube-apiserver-2-instance_e3c246ba-7357-4447...	kube-apiserv...	192.168.0.240	Running	●		0.5	1 GiB
	kube-apiserver-1-instance_0c909614-3594-46ae...	kube-apiserv...	192.168.0.242	Running	●		0.5	1 GiB
<input type="checkbox"/>	kubernetes.6d8ccbde-3cde-11e8-baa4-eec7567ec...	kubernetes	192.168.0.240	Running	●		1	1 GiB
	etcd-0-peer_75ac072b-7173-4a66-80c9-d3cb7b1...	etcd-0-peer	192.168.0.241	Running	●		0.5	1 GiB
	kube-scheduler-0-instance_2a713e7c-c550-4e16-...	kube-schedu...	192.168.0.241	Running	●		0.5	512 MiB
	kube-scheduler-2-instance_668b4987-538c-416...	kube-schedu...	192.168.0.240	Running	●		0.5	512 MiB
	kube-scheduler-1-instance_8d99309c-5a9d-44e7...	kube-schedu...	192.168.0.242	Running	●		0.5	512 MiB
	etcd-1-peer_145ec72e-8e18-4f3d-b6c2-f297900...	etcd-1-peer	192.168.0.240	Running	●		0.5	1 GiB
	etcd-2-peer_44202f18-78e7-4b92-9117-21ed4ba...	etcd-2-peer	192.168.0.242	Running	●		0.5	1 GiB

云基础设施

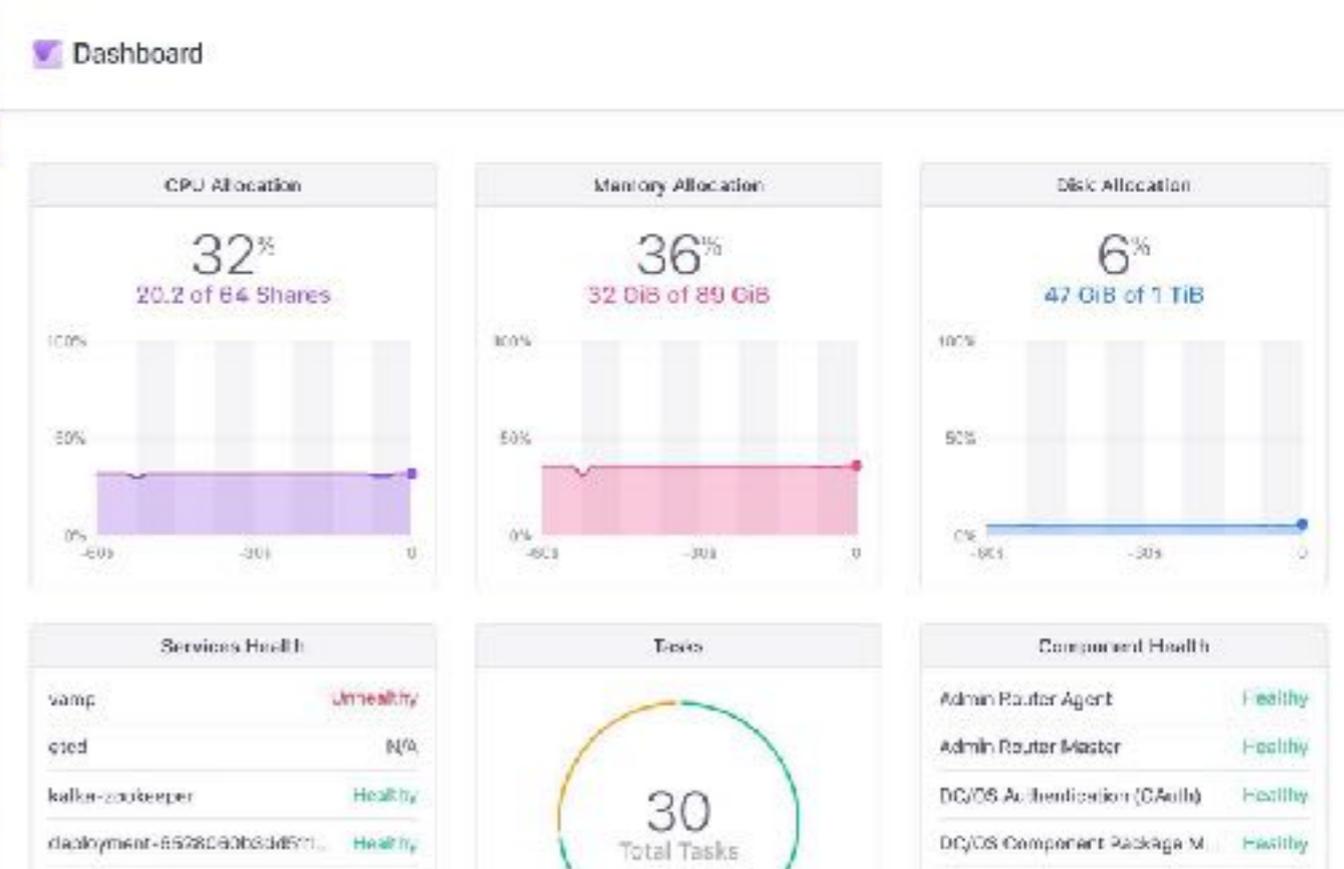
私有云/公有云/混合云

Azure/阿里云/AWS/Docket/Mesos/Rancher

自建、租用、多线多中心

私有容器云

5台二手DELL R710组成的私有容器云，80 Cores, 120G 内存，1.5T SAS硬盘，费用7800 RMB左右。基于DC/OS、Kubernetes和OpenShift，小团队和小公司可以自己弄一个私有容器云，用于研发、CI/CD和自动化。



DC/OS ON AZURE

mydcoscluster
Resource group

Settings Add Delete

Essentials

Subscription name: **Mesosphere Developer Subscription**
Subscription ID: **b1ed8f21-ceda-41d5-a0a8-95960b5340c2**
Last deployment: **4/15/2016 (Succeeded)**
Location: **West US**

Deployment history
mydcoscluster

mesosphere.dcos-previewdcos-2016031...
4/15/2016, 6:11:39 PM

All settings →

dcos-master

- Dashboard
- Services
- Jobs
- CI/CD
- RESOURCES
 - Nodes
 - Networking
- SYSTEM
 - Overview
 - Components
 - Settings
 - Organization

Dashboard

DPU Allocation
53%
12.75 of 24 Shares

Memory Allocation
36%
29 GiB of 82 GiB

Disk Allocation
21%
45 GiB of 208 GiB

Services Health

consul	N/A
kafka	N/A
marathon-lb	Healthy
traefik	Healthy
zeppelin	Healthy

VIEW ALL 6 SERVICES

Tasks
13 Total Tasks

13 Tasks running | 0 Tasks staging

Component Health

Admin Router Agent	Healthy
Admin Router Master	Healthy
Admin Router Follower	Healthy
Admin Router Follower	Healthy
Admin Router Follower Timer	Healthy

VIEW ALL 43 COMPONENTS

Nodes
6 Connected Nodes

RANCHER ON 青云

全部应用

添加应用

从应用商店添加

排序:

状态

名称

 + Demo	添加服务	3 服务	6 容器	 	
 + Demo-CICD	添加服务	3 服务	10 容器	 	
 + Demo-CIPipeline	添加服务	3 服务	3 容器	 	
 + demo-nginx	添加服务	2 服务	3 容器	 	
 - Demo-Prometheus	已经是最新版本 添加服务	9 服务	22 容器	 	
 Active	alertmanager ⓘ	镜像: prom/alertmanager 端口: 9093	服务-	1 容器	 
 Active	cadvisor ⓘ	镜像: google/cadvisor:latest	服务-	3 容器	 
 Active	grafana '+ 1 从容器' ⓘ	镜像: grafana/grafana:4.2.0 端口: 3000	服务-	2 容器	 
 Active	influxdb ⓘ	镜像: tutum/influxdb:0.10 端口: 2003	服务-	1 容器	 
 Active	node-exporter ⓘ	镜像: prom/node-exporter:latest	服务-	3 容器	 

WHY RANCHER

Rancher的核心价值——简单即美

企业容器云

应用服务目录	Helm, ...
编排	Compose, Kubernetes, Marathon,
调度	Swarm, Kubernetes, Mesos, ...
监控	cAdvisor, Prometheus, Datadog, ...
权限访问控制	LDAP, AD, GitHub, ...
镜像库	Nexus, Artifactory, DTR...
容器引擎	Docker, runC, Rocket ...
安全	Notary, Vault, ...
网络	Cisco, VXLAN, IPSEC, HAProxy, ...
存储	Ceph, Gluster, Swift, ...
分布式数据库	Etd, Consul, MongoDB, ...

OR

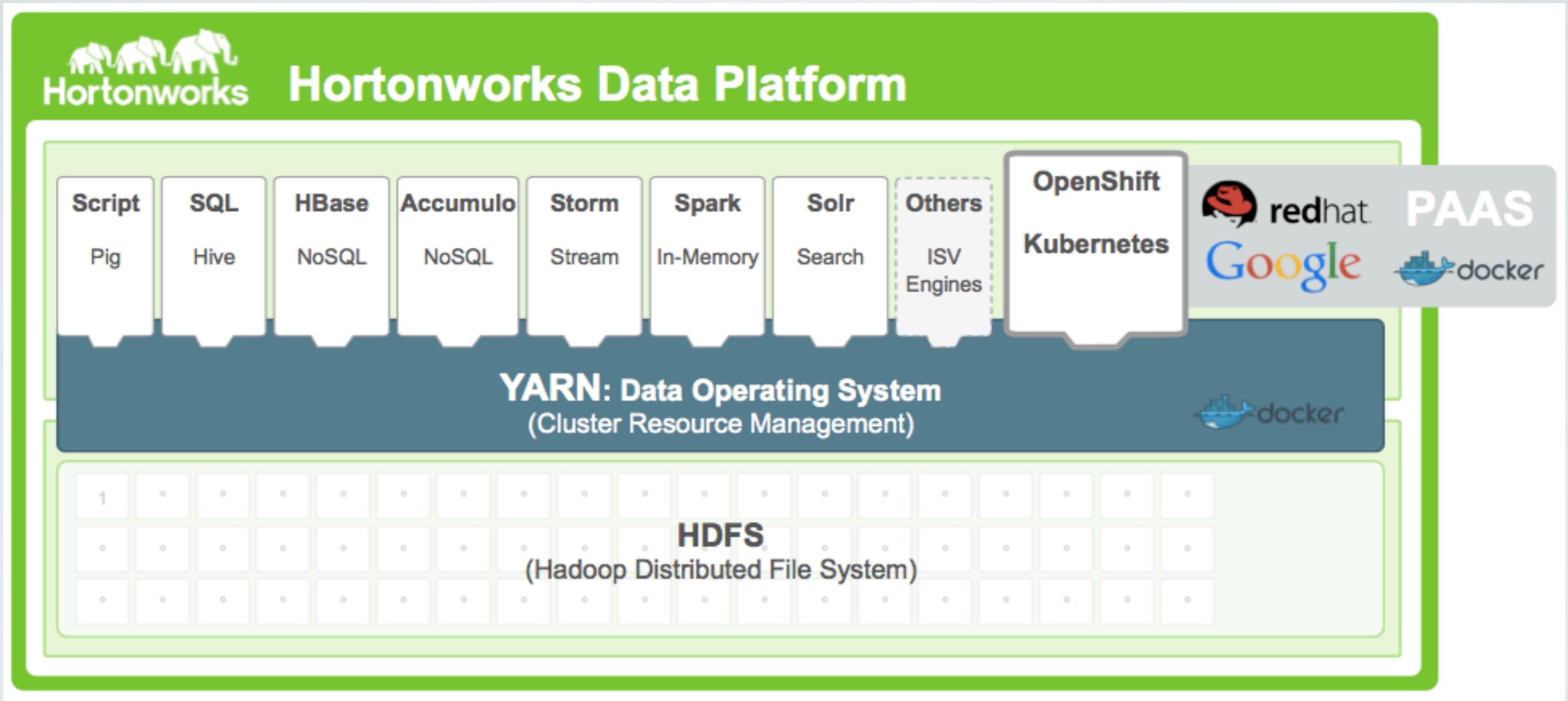


全栈化

简单易用

企业级K8S容器平台

YARN: BIG DATA AND PAAS TOGETHER



开源治理、企业研发

大教堂与市集，The Apache Way

Linux基金会、Apache基金会

企业的研发管理，更好的结合开源软件治理和开源生态系统建设

Open Source Office – Non Development Activities



Relationships Collaborations

Linux Foundation

Open Invention Network

Software Freedom Law Center

Mozilla Foundation

GNOME Foundation

Free Software Foundation

Apache Foundation

Software Conservancy Center

Open Source Legal Compliance

Manage and support open source compliance process

Advise on open source compliance

Help resolve compliance inquiries

Provide guidelines and checklists to enable internal education and awareness on the topic

Sensing on emerging open source legal concerns

Strategy, Community, Engagement

Develop, maintain and execute corporate open source strategy & policy

Internal and external evangelism

Represent open source views in internal policy discussions

Launch new open source projects, support open sourcing internal code

Sponsor, publish, speak at open source conferences

Organize community events

Develop and deliver internal open source training

Provide engineering mentorships to grown internal open source expertize

Conduct internal technical workshops

Advise on community matters



TODO

<https://todogroup.org>

开源指南、开源项目服务

最大限度优化组织中运行开源计划或启动开源项目的实践

创建可持续的开源项目生态系统

DEVOPS

DEVOPS驱动组织高效协作和技术转型

DevOps如何驱动?

研发、运营、服务

DEVOPS

DEVOPS是覆盖IT全生命周期的生产线

SRE是DEVOPS模型的最佳实践

DevOps整个平台划分成了三大领域：敏捷过程、持续交付、持续改进。

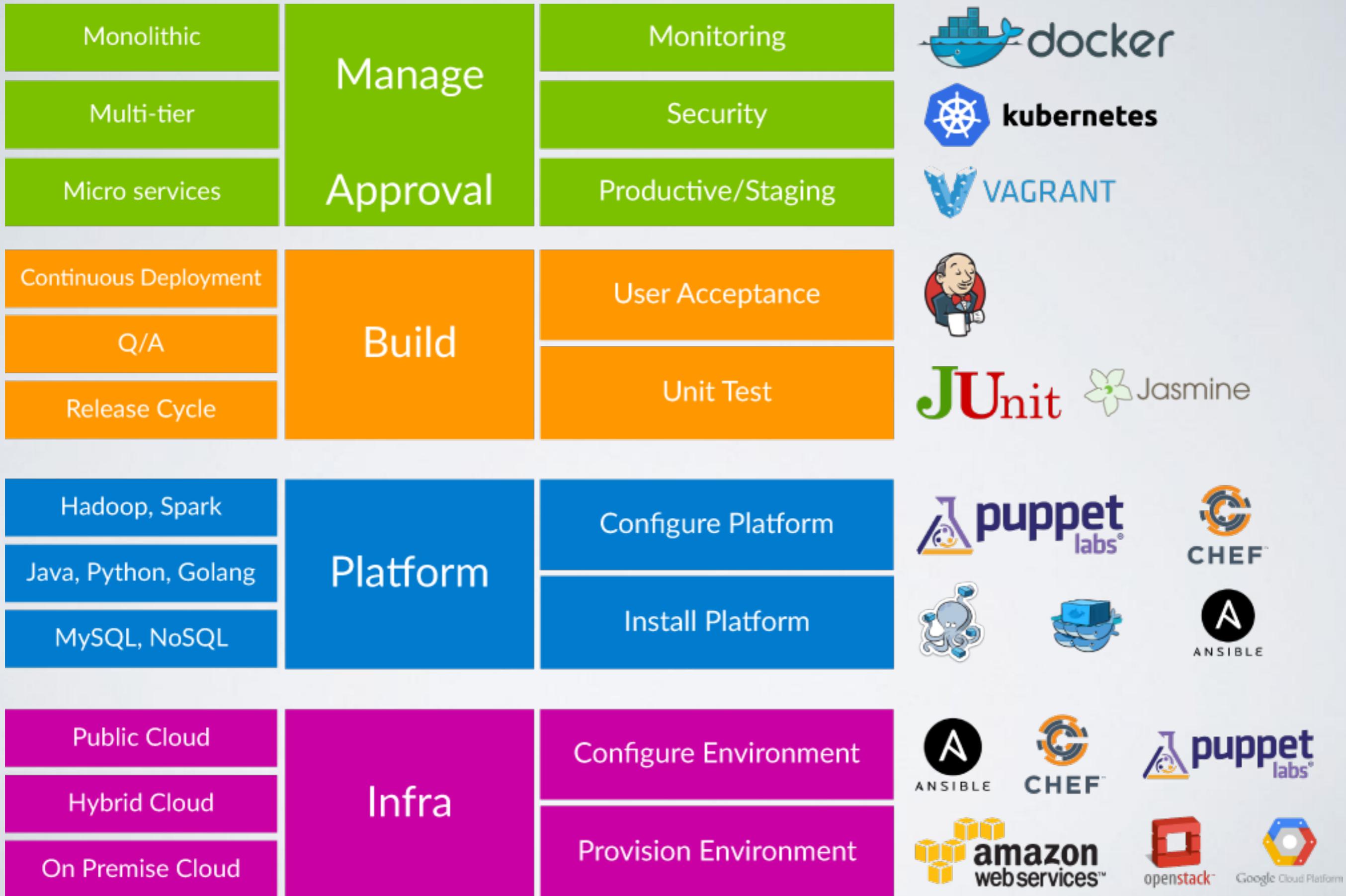
敏捷过程：包含产品管理、项目管理、任务管理、进度管理、计划管理等，覆盖产品、项目的全生命周期。

持续交付：包含代码库的管理、持续集成、部署、交付流水线等能力，打通从代码提交到部署上线的全流程。持续集成支持编译、打包、测试、工具四类构建任务，支持代码提交时触发构建、定时构建、手动构建三种构建触发策略。在部署方面，支持Android, iOS应用、传统war包、Java站点、Cassandra部署，支持灰度发布、滚动升级、蓝绿发布等多种部署策略。

持续改进：包含质量标准、质量监控、以及产品项目全生命周期过程中的各种度量报表，支撑企业的精益度量和统计分析。

支撑软件：GitLab、Jenkins、SMACK、Teambition、Phabricator(通过API将GitLab、Jenkins、SMACK相关数据进行聚合分析)、监控系统。

DEVOPS参考堆栈



研发

PIPELINES

BaaS、PaaS、PM、CI/CD

运营

支撑公司业务运营的ERP和SAAS服务

好的产品、应用、服务是运营出来的

SaaS、业务沉淀、模块化、应用市场

服务

ERP(CRM模块) + 呼叫中心

全渠道

网站社区、微信、微博、车友会、VIP用户

自动化

规范、适应和进化

1、精益看板和项目协作 2、CI/CD/DevOps自动化 3、APM监控基础设施

实施步骤

产品服务、研发运营、IT基础架构的持续改善和进化

1、提升PM/CI/CD/DevOps流程和效率 2、PaaS/BaaS/API 3、私有公有混合容器云

工程师&开源文化

黑客、创客、宽客、HACKATHON

精善语言、解决问题、参与开源、代码贡献(Committer)

Linux, Apache, FSF, Mozilla ...

开放式组织

面向未来的组织管理新范式

开放式管理和混序管理是互联网时代最具代表性的组织方式和管理模式

做什么(确定方向)、为什么(激励和启发)、怎么做(完成任务)

依托社区、网聚人的力量...

布道师

布道之道，引领团队拥抱技术创新

传达理念、分享经验、展示技术、建立信任



星主：龙辉

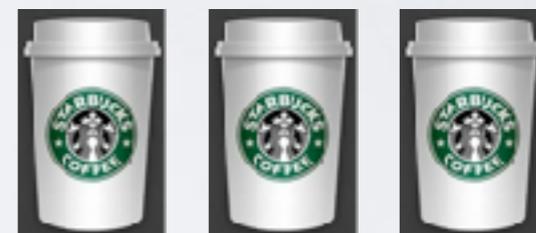
灰狐的朋友们

 知识星球

微信扫描预览星球详情



谢谢



技术交流、咨询培训、支持服务