

K8S 1.18版本 (<https://www.kubernetes.org.cn/tags/kubernetes-1-18>) 1.17 (<https://www.kubernetes.org.cn/tags/kubernetes-1-17>) 1.16 (<https://www.kubernetes.org.cn/tags/kubernetes-1-16>)

CI/CD (<https://www.kubernetes.org.cn/tags/cicd>) 网络 (<https://www.kubernetes.org.cn/tags/网络>) 存储 (<https://www.kubernetes.org.cn/tags/存储>) 安全 (<https://www.kubernetes.org.cn/tags/安全>)

监控 (<https://www.kubernetes.org.cn/tags/监控>) 运维 (<https://www.kubernetes.org.cn/tags/运维>) 认证 (<https://www.kubernetes.org.cn/tags/认证>) 登录 (<https://www.kubernetes.org.cn/wp-login.php?cms=k8s>) | 加入我们 (<https://www.kubernetes.org.cn/%e8%81%94%e7%b3%bb%e6%88%91%e4%bb%ac>)

开发实践 (<https://www.kubernetes.org.cn/practice>) 行业动态 (<https://www.kubernetes.org.cn/news>) 入门教程 (<https://www.kubernetes.org.cn/course>)

安装教程 (<https://www.kubernetes.org.cn/course/install>) 文档下载 (<https://www.kubernetes.org.cn/%E6%96%87%E6%A1%A3%E4%B8%8B%E8%BD%BD>)

QQ/微信群 (<https://www.kubernetes.org.cn/kubernetes>) 视频 (<https://www.kubernetes.org.cn/video>) 活动 (<https://www.kubernetes.org.cn/meetup>)

中文文档 (<http://docs.kubernetes.org.cn>)

Kubernetes 中文社区

Q

Kubernetes CKA实战培训 突击CKA全球认证

官方认证CKA讲师、实操环境实战、现场答疑互动

开班城市：北京 / 上海 / 深圳 / 武汉 / 成都 / 杭州 以及线上直播班

点击查看

(<https://www.kubernetes.org.cn/peixun>)

您目前处于：社区首页 (<https://www.kubernetes.org.cn>) > DevOps (<https://www.kubernetes.org.cn/devops>) > 阿里巴巴智能化代码平台的探索与实践

# 阿里巴巴智能化代码平台的探索与实践 (<https://www.kubernetes.org.cn/7869.html>)

2020-07-08 22:53 Yvonne (<https://www.kubernetes.org.cn/author/yvonne>) 分类：DevOps (<https://www.kubernetes.org.cn/devops>) 阅读(294) 评论(0)

2020年6月13日，在2020阿里巴巴研发效能峰会“架构设计与代码智能专场”中阿里巴巴高级技术专家张玉明（玄坛）发表题为《阿里巴巴智能化代码平台的探索与实践》的主题演讲，详细介绍了阿里巴巴打造智能化代码管理平台的起因、技术难点和解决思路以及应用案例，并发布云效智能研发助手“云豆”。本文节选自玄坛的分享，为方便开发者阅读，删除“背景”介绍部分，将集中讲解云效代码管理平台中智能化技术的应用以及实现原理。

## 人工智能赋能开发者 解决软件研发问题

作为一名开发者，在日常工作中会遇到哪些问题？主要有以下四点：需求不明确或需求经常变动；缺陷难发现：在开发的代码中经常隐藏难以发现的BUG；编码效率低：编码效率不够高，没办法合理安排编码时间；线上问题难定位。

有没有解决以上问题的灵丹妙药呢？云效团队给出的解法是“工具+数据+智能”。其实我们每天在使用的云效代码管理平台、钉钉等都是非常好用的工具。在这些工具基础之上，结合大数据、智能化手段是否可以在代码编写提效、代码质量、安全提升方面做一些探索呢？



首先看一下我们在“大数据”方面是如何做的。在阿里巴巴集团内部的代码管理平台上每天都会产生大量代码数据，我们将其中公开的代码离线存储到代码仓库中去构建代码数仓。基于代码数仓可以构建代码图谱，比如我们会以代码、用户、代码库这些维度进行抽象，将其转化为实体，并通过对这些实体关系的标

## 热门推荐

- 你的微服务还差个容错机制

2020-03-26 (<https://www.kubernetes.org.cn/7060>)
- Kubernetes 1.18GA，15个稳定11个beta，引入kubectrl debug命令

2020-03-26 (<https://www.kubernetes.org.cn/7055>)
- 英国Monzo银行，用K8s管理1600个微服务实践

2020-03-21 (<https://www.kubernetes.org.cn/7001>)
- Java vs. Go 微服务 – 负载测试(复赛)

2020-03-20 (<https://www.kubernetes.org.cn/6988>)
- 与时俱进 – 为什么要使用云原生数据库？

2020-03-15 (<https://www.kubernetes.org.cn/6953>)
- 2020年Service Mesh 三大发展方向

2019-12-13 (<https://www.kubernetes.org.cn/6255>)

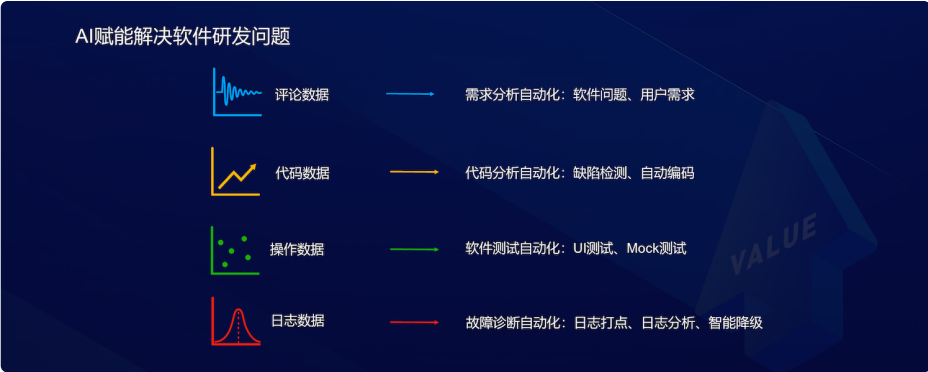
## 社区标签

- BoCloud博云 (<https://www.kubernetes.org.cn/tags/bocloud%e5%8d%9a%e>)

CI/CD (<https://www.kubernetes.org.cn/tags/cicd>)

CNCF (<https://www.kubernetes.org.cn/tags/cncf>)

签化，从而构建代码画像、用户画像、代码库画像。进而为上层的智能化服务提供有力的数据支撑。同时我们以代码为中心，去拉通研发数据、协作数据、应用运行数据，去构建研发图谱，围绕整个研发生命周期去提供数据方面的支撑。



那么有了这些数据，我们可以做什么事情？基于评论数据，结合智能化手段，可以做到需求分析自动化，比如可以基于前台用户咨询的问题，结合智能化手段来帮助产品经理明确用户诉求，提炼产品文档。基于代码数据，可以实现代码分析自动化，如可以应用到代码缺陷检测和代码自动编写等领域中。基于用户的操作数据，可以做到软件测试自动化，如现在很火爆的UI测试、Mock测试等。基于日志数据，可以实现故障诊断自动化，如日志自动打点、日志智能分析以及服务智能降级等。

云效推智能研发助手“云豆”代码评审更智能

接下来看一下我们是如何在云效代码管理平台（Codeup）中将代码智能化能力进行落地的。首先来看代码评审。



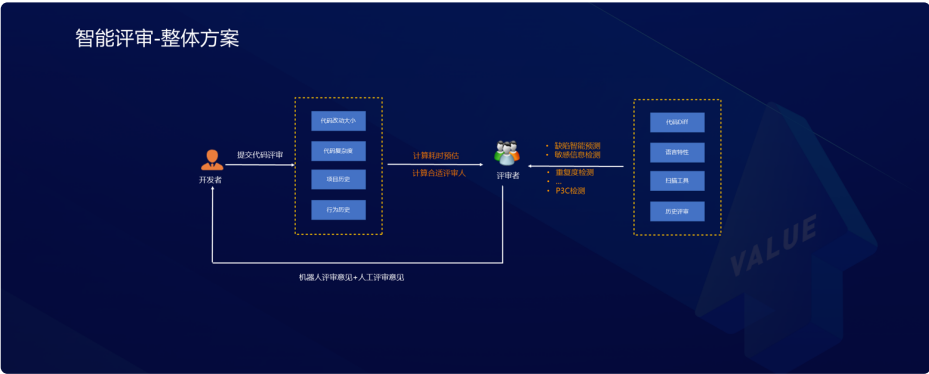
传统的代码评审分为三个阶段：新建评审、评审中、评审结束。在这三个阶段中，用户都会产生哪些行为呢？在新建评审阶段，首先需要选定某个分支，然后选择评审人，最后需要填写评审描述，通过这三步操作就新建了一个评审任务。在评审中，评审者首先要查看评审文件，然后提出问题，最后由开发者解决问题。如果评审通过，并完成合并代码，则评审结束。

在上述过程中存在哪些痛点呢？在选择评审人环节，开发者经常不知道应该选择谁作为这段代码的评审者；在评审环节，评审者写不清评审描述；变更文件过多，无法有效评审；在评审结束环节，经常需要人工去解决代码冲突这类问题。

其实有“痛点”就有“机会点”，针对这些在代码评审过程中的痛点，我们还可以做什么呢？比如，是否可以智能推荐评审人？智能生成评审描述？是否可以对改动过多的“大评审”进行自动化拆解？在出现代码冲突时，是否可以智能化修复？

- DevOps (https://www.kubernetes.org.cn/tags/
- Docker (https://www.kubernetes.org.cn/tags/
- etcd (https://www.kubernetes.org.cn/tags/etcd)
- GO (https://www.kubernetes.org.cn/tags/go)
- Helm (https://www.kubernetes.org.cn/tags/he
- Istio (https://www.kubernetes.org.cn/tags/istio)
- Jenkins (https://www.kubernetes.org.cn/tags/jenk
- k8s代码解读 (https://www.kubernetes.org.cn/tags/k8s%e4%bb%a3%e
- kubeadm (https://www.kubernetes.org.cn/tags/kube
- kubernetes (https://www.kubernetes.org.cn/tags/kuberenc
- Kubernetes (https://www.kubernetes.org.c
- Kubernetes 1.4 (https://www.kubernetes.org.cn/tags/kubernetes 1-4)
- Kubernetes 1.5 (https://www.kubernetes.org.cn/tags/kubernetes 1-5)
- Kubernetes 1.6 (https://www.kubernetes.org.cn/tags/kubernetes 1-6)
- Kubernetes 1.7 (https://www.kubernetes.org.cn/tags/kubernetes 7)
- Kubernetes 1.8 (https://www.kubernetes.org.cn/tags/kubernetes 1-8)
- Kubernetes 1.9 (https://www.kubernetes.org.cn/tags/kubernetes 1-9)
- Kubernetes 1.10 (https://www.kubernetes.org.cn/tags/kubernetes 1-10)
- Kubernetes 1.13 (https://www.kubernetes.org.cn/tags/kubernetes 1-13)
- OpenStack (https://www.kubernetes.org.cn/tags/openstack)
- PaaS (https://www.kubernetes.org.cn/tags/paas)
- Pod (https://www.kubernetes.org.cn/tags/pod)
- Prometheus (https://www.kubernetes.org.cn/tags/prometheus)
- Rainbond (https://www.kubernetes.org.cn/tags/rainbond)
- Rancher (https://www.kubernetes.org.cn/tags/rancher)
- Serverless (https://www.kubernetes.org.cn/tags/server
- Service (https://www.kubernetes.org.cn/tags/service)

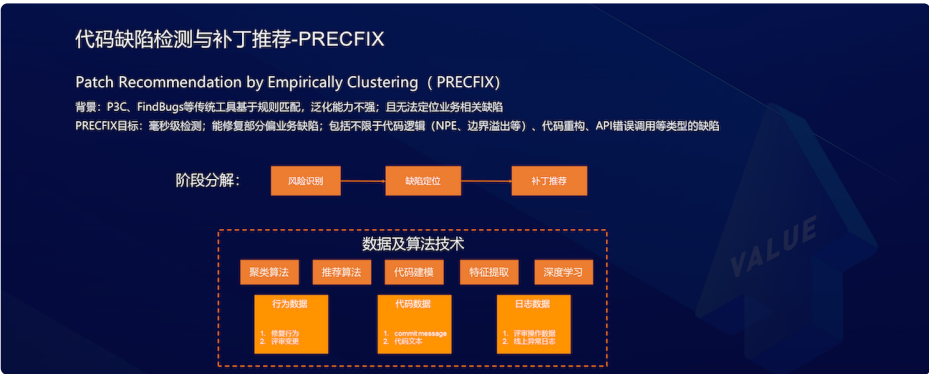
“云豆”可以在智能引导环节分享更规范的工作方法，在智能评审环节快速定位问题高效完成工作，在智能答疑环节推荐最优解法。比如在新建评审环节，当选择好评审分支后，云豆可以智能推荐最佳评审者；云豆支持预估评审耗时，并可以根据预估评审耗时合理安排评审优先级，帮助开发者利用好碎片时间评审。在代码质量方面，云豆具备自动化质量和安全检测能力，并可自动推荐修复方案。



我们可以将云效智能研发助手“云豆”提供的智能化评审过程抽象为上面这张图，当一个开发者新建了一个代码评审，云豆可以结合智能化及传统自动化检测手段对这段代码进行全方位的检测，并给出一个非常综合的评审意见。此时评审者只需要在云豆提供的评审意见之上，结合自己的工作经验及对业务逻辑的理解给出补充的评审意见即可。

智能研发助手云豆背后的能力：代码缺陷检测与补丁推荐-PRECFIX

在云效智能研发助手云豆提供的智能评审过程中，究竟涉及哪些智能化的能力呢？首先，我们来看代码缺陷检测与补丁推荐技术-PRECFIX。



在日常研发工作中，我们经常会用到P3C、PMD、FindBugs等传统自动化检测工具，但这些工具并不能完全解决阿里巴巴面临的代码质量问题。因为传统工具多是基于规则匹配，泛化能力不强，基于业务场景的缺陷很难识别出来。因此我们希望有一种对缺陷类型泛化能力比较强的缺陷检测方法或者工具，于是提出PRECFIX方法（Patch Recommendation by Empirically Clustering）。PRECFIX的目标是通过智能化手段，可以实现毫秒级检测，能够修复部分偏业务缺陷，并给出对应的补丁推荐方案。

PRECFIX代码检测过程主要分为三个阶段：风险识别、缺陷定位、补丁推荐。这个过程会涉及行为数据、代码数据、日志数据的处理，以及聚类算法、推荐算法、代码建模、特征提取、深度学习等算法技术。

service mesh  
(<https://www.kubernetes.org.cn/tags/service-mesh>)

云原生  
(<https://www.kubernetes.org.cn/tags/>

企业案例  
(<https://www.kubernetes.org.cn/tags/%e4>

存储  
(<https://www.kubernetes.org.cn/tags/%e5%ad%91>

安全  
(<https://www.kubernetes.org.cn/tags/%e5%ae%89%97>

容器  
(<https://www.kubernetes.org.cn/tags/%e4%b9%99>

容器云  
(<https://www.kubernetes.org.cn/tags/%e5%ae%b9%e5%99%9d>

容器云平台  
(<https://www.kubernetes.org.cn/tags/%e5%ae%b9%e5%99%a8%e4%ba%91>

微服务  
(<https://www.kubernetes.org.cn/tags/%e4%b9%99>

日志  
(<https://www.kubernetes.org.cn/tags/%e6%97%a5%e5%bf%9c>

灵雀云  
(<https://www.kubernetes.org.cn/tags/%e7%81%b5%e9%9b%9c>

灵雀云 Jenkins  
(<https://www.kubernetes.org.cn/tags/%e7%81%b5%e9%9b%9c>

监控  
(<https://www.kubernetes.org.cn/tags/%e7%9d%9c>

网络  
(<https://www.kubernetes.org.cn/tags/%e7%9d%9c>

运维  
(<https://www.kubernetes.org.cn/tags/%e8%bf%90%e7%bb%9c>

最新评论

loong576 3天前说：  
master01初始化后就启动了k8s的服务。master02和master03加入之前先保证master01的k8s服务正常，可通过下面三种方式检查：1.'ss -nltlgrep 6443'查看6  
(<https://www.kubernetes.org.cn/6632.html#comment-1480>)

zhangsn 3天前说：  
我也遇到同样问题，上面安装步骤中并未提及如何启动VIP节点的k8s的服务，所以master2和3始终加入不了集群  
(<https://www.kubernetes.org.cn/6632.html#comment-1479>)

loong576 3天前说：  
文章‘第七部分4. 安装flannel网络’有下载链接，或者直接文来github上下载： <https://github.com/loong576/Centos7.6-install-k8s-v1.16>. (<https://www.kubernetes.org.cn/6632.html#comment-1478>)

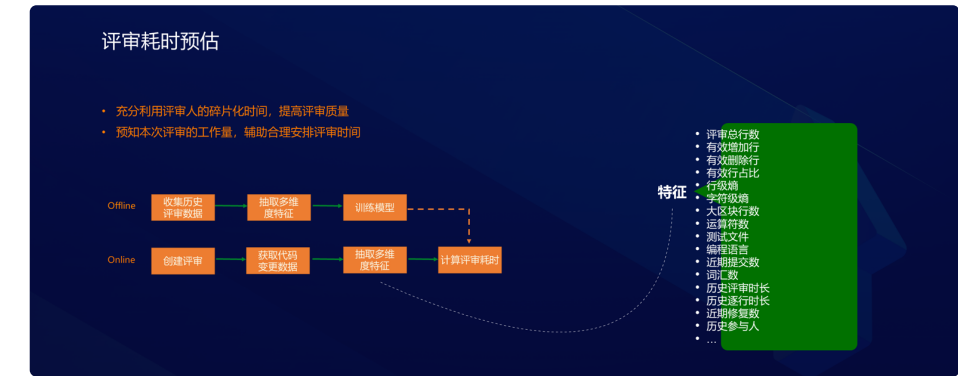
xiao 4天前说：  
在的kubernetes.com下载地址不对，是k8s地址应该对。

PRECIFIX的技术实现思路其实并不复杂，首先我们从阿里巴巴集团内部海量的公开代码数据中提取代码提交（Commit）信息，基于一定规则挖掘出“缺陷修复对”（缺陷代码片段和补丁代码片段的组合），然后通过聚类算法将相似的“缺陷修复对”聚类，提取出对应的代码模板。在用户发起评审时，系统会自动扫描变更代码，若与缺陷模板库匹配，则会推荐给用户补丁模板用于修复。



PRECIFIX方法已经在阿里巴巴集团内部落地，其误报率在5%以下，在内部公开代码库中扫描出了800多种缺陷类型，3万多个缺陷，提取出了3000多个补丁模板。

智能研发助手云豆背后的能力：评审耗时预估



当存在多个待处理评审的情况时，“评审耗时预估”可以帮助评审人预知评审的工作量，合理安排评审时间，利用碎片化时间进行充分的评审，提升评审效率。

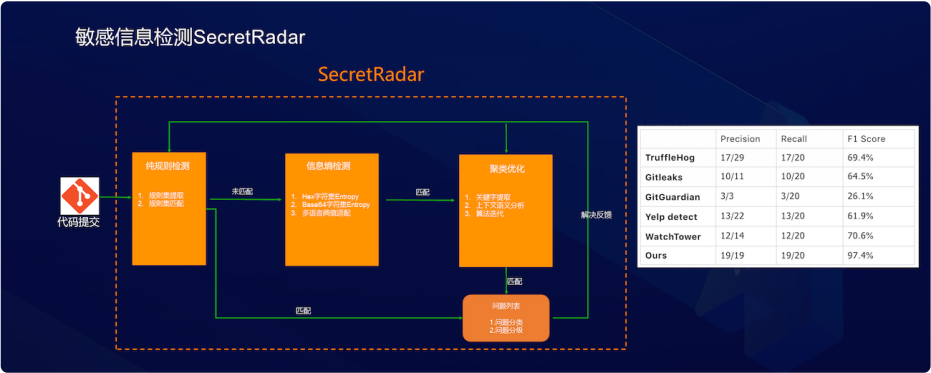
“评审耗时预估”的技术实现思路是怎样的呢？首先从几百万次的代码评审历史浏览数据中抽取特征，包括评审总行数、编程语言、历史评审时长、历史逐行时长等几十个维度的特征，从而去训练机器学习模型。当一个开发者提交了代码评审时，“云豆”会从他提交的diff文件中提取这些特征，从而估算这次评审需要的耗时。

智能研发助手云豆背后的能力：敏感信息检测SecretRadar





近年来，业内发生多起敏感信息（API Key、Database credential、Private token）通过某些站点被无意识地泄露出去的事件，给企业带来了安全风险。开发者和企业管理者亟需一款稳定健全的敏感信息检测方法和系统。通过调研我们了解到，目前已有的敏感信息检工具要么单纯考虑规则匹配，要么采用信息熵技术实现，其召回率或准确率均无法满足预期。因此我们在规则匹配和信息熵技术的基础上，结合了多层检测模型和上下文语义检测，打造了一款敏感信息检测工具——SecretRadar。



SecretRadar的技术实现思路主要分为三层，第一层采用规则匹配这种传统敏感信息识别技术，“规则匹配”具有良好的准确度和扩展性，但是非常依赖比较固化的长度、前缀、变量名，难以应对不同开发者的不同编码风格，容易造成漏报。针对难以固定规则捕捉的场景，在第二层我们采用了信息熵算法。信息熵算法用于衡量代码行混乱程度，对随机生成型密钥和随机身份信息识别效果良好。但信息熵算法也有其局限性，伴随召回的提升是误报率的增加。因此在第三层我们采用了模板聚类的方法进行过滤优化。针对信息熵结果集聚合提取常见关键字，并结合上下文分析完成二次过滤。同时通过问题的修复情况建立二分类数据集，完成算法优化。进而从词法识别迭代为语义识别。

智能研发助手云豆背后的能力：代码补全

代码补全是当前代码智能化领域研究的热点。目前代码补全工具主要分为三种：基于语法解析器的代码补全；基于统计机器学习的代码补全；基于深度学习的代码补全。基于语法解析器的代码补全工具最为大家熟知，我们平时使用的IDE中内置的补全工具基本都属于这类。这种补全工具有个问题，就是补全结果没有进行合理排序。基于统计机器学习的代码补全工具不仅可以给出更加合理的补全候选方案，还可以从“补全概率”的维度对候选方案进行排序。但这种补全方式也有其局限性，推荐的结果比较固定。

最近几年，随着NLP（自然语言处理）领域对自然语言建模能力的大幅度提升，这种能力也被应用到代码补全上。相对于基于统计机器学习的代码补全工具而言，基于深度学习的代码补全系统能够更大范围的了解上下文语义，并结合大量数据，给开发者提供动态的代码补全推荐。当然这种方式在训练阶段以及推理计算时会需要更强大的算力，因此在普通的PC机上基本没办法使用这种系统。业内通常的做法是将训练好的模型部署到远端服务器上，本地去进行实时查询，因此对网络的依赖比较强。

在代码补全的探索中我们遇到很多挑战和难点，主要有四个方面：不完整代码上下文理解难度大；代码片段补全准确率要求极高，对错误的容忍度非常低；多Token补全搜索空间无限大，算法效率和准确度难以提升；代码中单词的数目是无限大的，一般模型难以处理如此大的词表。

针对以上问题，我们的策略是采用“算法手段+工程手段”相结合的解决方案。算法方面，采用本地+深度模型共同补全，挑选合适的分词算法（如BPE分词算法），并结合规则匹配的手段进行实现。在工程方面，通过整合现有语法服务给用户带来综合体验的提升。



我们的代码补全流程设计如图所示，分为预处理阶段、实时补全阶段、后处理阶段。在预处理阶段，会进行变量/字符串替换、分词、索引变量等处理。在实时补全阶段，采取本地个性化补全模型和云端深度补全模型相结合的方法，然后对多模型候选项进行合并、筛选和重排。在后处理阶段，会进行变量名替换和填充、语义格式化等操作，并将最终补全候选方案推荐给开发者。

总结：

目前云效智能研发助手“云豆”主要可提供智能代码评审、评审耗时预估、智能推荐评审人、代码补全等智能化能力，未来“云豆”还可以结合开发者个人研发习惯，提供更多功能。云豆已经集成在云效代码管理平台（Codeup）中，大家可以登录云效官网免费“领养”。

以上内容节选自张玉明（玄坛）的主题演讲《阿里巴巴智能化代码平台的探索与实践》，如果想了解完整分享内容可以进入2020阿里巴巴研发效能峰会官方网站收看回放视频。

了解更多信息登录：<https://www.aliyun.com/product/yunxiao>



关注微信公众号，加入社区



(<http://www.aliyun.com/product/yunxiao>)

上一篇：厉害了，云效自研代码缺陷检测技术Prefix被软件工程顶会录用啦！  
(<https://www.kubernetes.org.cn/7984.html>)

Kubernetes 1.18 版本

我们很高兴宣布Kubernetes 1.18的交付。这是我们2020年的第一版！Kubernetes 1.18包含38个增强功能：其中15个功能已趋于稳定，beta版本中有11个，alpha版本中有12个。  
(<https://www.kubernetes.org.cn/tags/kubernetes-1-18>)



关注「K8S中文社区」微信公众号  
回复“文档”  
获取K8S文档下载链接  
回复“加群”  
加入K8S微信技术交流群

最新文章



Docker私有镜像拉取错误no basic auth credentials  
2020-07-11 评论()  
(<https://www.kubernetes.org.cn/7994>)



技术干货分享！Calico IPAM源码解析  
2020-07-09 评论()  
(<https://www.kubernetes.org.cn/7989>)



阿里巴巴智能化代码平台的探索与实践  
2020-07-08 评论()  
(<https://www.kubernetes.org.cn/7869>)



厉害了，云效自研代码缺陷检测技术Prefix被软件工程顶会录用啦！  
2020-07-08 评论()  
(<https://www.kubernetes.org.cn/7984>)

Kubernetes 版本资讯

- Kubernetes v1.18 正式版已发布  
(<https://www.kubernetes.org.cn/7055.html>)
- Kubernetes v1.17 正式版已发布  
(<https://github.com/kubernetes/kubernetes/releases/tag/v1.17.0>)
- Kubernetes v1.16 正式版已发布  
(<https://www.kubernetes.org.cn/5838.html>)
- Kubernetes v1.15 正式版已发布  
(<https://www.kubernetes.org.cn/tags/kubernetes-1-15>)
- Kubernetes v1.14 正式版已发布  
(<https://www.kubernetes.org.cn/5204.html>)



## 从这里开启您的 Kubernetes 之旅

100% 开源
Web UI
离线部署
一键扩容
应用商店

(<https://github.com/KubeOperator/KubeOperator>)

### 相关推荐

- 厉害了，云效自研代码缺陷检测技术Precfix被软件工程顶会录用啦！(<https://www.kubernetes.org.cn/7984.html>)
- 重大福利，JFrog发布面向社区的免费安全的HelmChart中央存储库ChartCenter(<https://www.kubernetes.org.cn/7976.html>)
- 切忌一步到位，谈谈DevOps实施落地(<https://www.kubernetes.org.cn/7966.html>)
- 云原生成为工业互联网新引擎 共同赋能制造业转型升级(<https://www.kubernetes.org.cn/7963.html>)
- DevSecOps -10项最佳安全实践(<https://www.kubernetes.org.cn/7961.html>)
- RPM索引在Artifactory中是如何工作(<https://www.kubernetes.org.cn/7959.html>)
- 5种主要的软件架构模式(<https://www.kubernetes.org.cn/7863.html>)
- “2020行业云原生应用报告指南”发布，传统企业面临的又一次关键技术抉择(<https://www.kubernetes.org.cn/7862.html>)

### 评论

社区交流

2 5 3 3 3

提交评论

昵称

昵称 (必填)

邮箱

邮箱 (必填)

网址

网址