

本科生《数据库系统原理》

系统规划与可行性分析报告

|  |  |
| --- | --- |
| **姓 名：** | **文楷昂** |
| **学 号：** | **19281141** |
| **班 级：** | **AI1901** |
| **日 期：** | **2021年1月5日** |

目录

[1. 系统名称 3](#_Toc13749)

[2. 组织架构 3](#_Toc29351)

[3. 技术架构 3](#_Toc28434)

[4. 调研分析业务场景 4](#_Toc31994)

[5. 系统性能指标 6](#_Toc13048)

[6. 系统战略 6](#_Toc30285)

[7. 投资和运营成本 7](#_Toc12054)

[8. 技术可行性 7](#_Toc9294)

[9. 专家论证 7](#_Toc11801)

# 系统名称

解惑论坛

# 组织架构

由互联网上公示的信息，知乎的组织架构如下：



其中标蓝的为我们设计的系统可能涉及到的部门。

# 技术架构

对知乎现技术架构进行调查，得到：

1．微服务架构，知乎从 11 年就开始了微服务的探索，尝试过 protocol buffers、Avro、Thrift，最终在 16 年确认使用 Thrift，同时使用 Consul 和 HAProxy 作为注册中心和负载均衡。是在 14 年确认的这套微服务架构，并且稳定使用到了现在。所以大家不要问为什么不使用 gRPC 了。

2．云平台，知乎有自己的内部研发的 ZAE ，绝大部分的在线业务容器在 15 年就已经全部跑在了 Docker 里，现在我们 HBase 和 Kafka 也是跑在容器里的。我们最开始使用的是Mesos 做的资源调度，现在已经切换到了 Kubernetes 。

3．部署平台，知乎的部署平台是与 ZAE 在一起的， 基于 Jenkins 搭建的自动集成，在 MR（Gitlab） 阶段自动使用 SonarQube 进行静态代码检查。部署分为测试环境，办公室环境，金丝雀1（灰度单个容器），金丝雀2（灰度 20% 流量），生产环境（100% 流量上线）。如果金丝雀阶段出现错误，会自动进行回滚操作。

4．监控，我们主要基于 Grafana、OpenTracing、Graphite 等搭建了监控系统。同时自研了 Halo 可以方便的是业务方观测到服务之间的依赖关系、响应时间（P95, P99, P999）、错误数。同时也进行了新技术的尝试，目前在业务容器监控使用了Prometheus 。

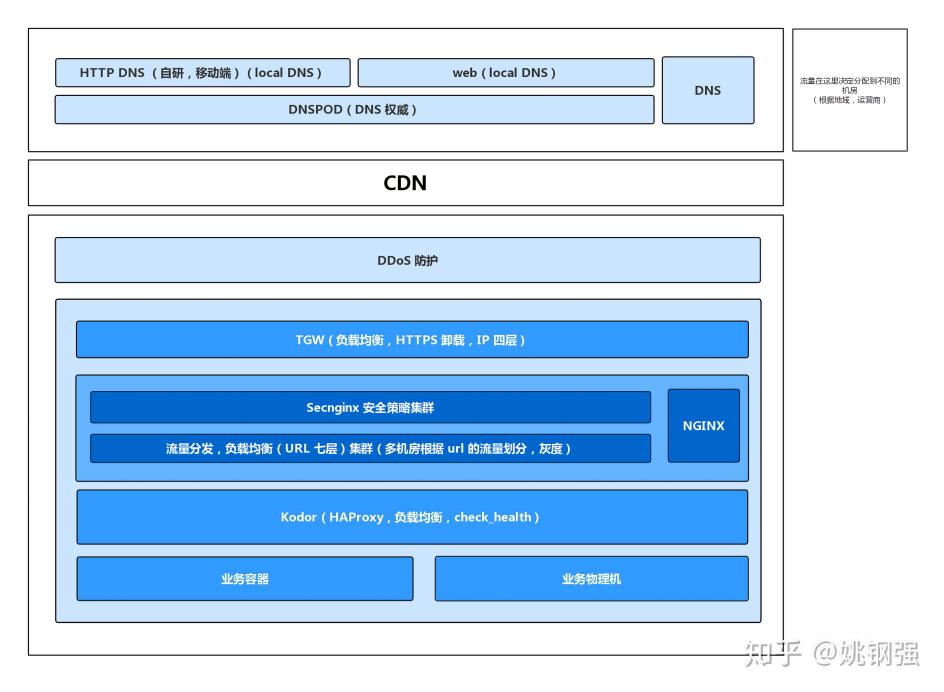
5．存储，主要是 MySQL、Redis、HBase；正在调研 TiDB，目前有一套生产集群上线准备给「已读」服务使用。

6．消息队列，早期使用自研的 Sink，目前使用 Kafka，同时提供在 Kafka 的基础上包装了Beanstalkd 作为任务队列方便业务进行使用。

7．编程语言，Python、Golang、Java、Rust。目前 Golang 使用比较广泛，Python 使用场景逐渐变少。Java 在一些算法项目和商业系统中有使用。搜索系统使用的 Rust 重写 Lucene，现并在此基础上重写了类 ES 的集群化搜索引擎。架构师 @孙晓光 会去 https://rustcon.asia/ 做分享。

8．目前已经搭建好完备的测试环境，会在 Q2 投入使用。

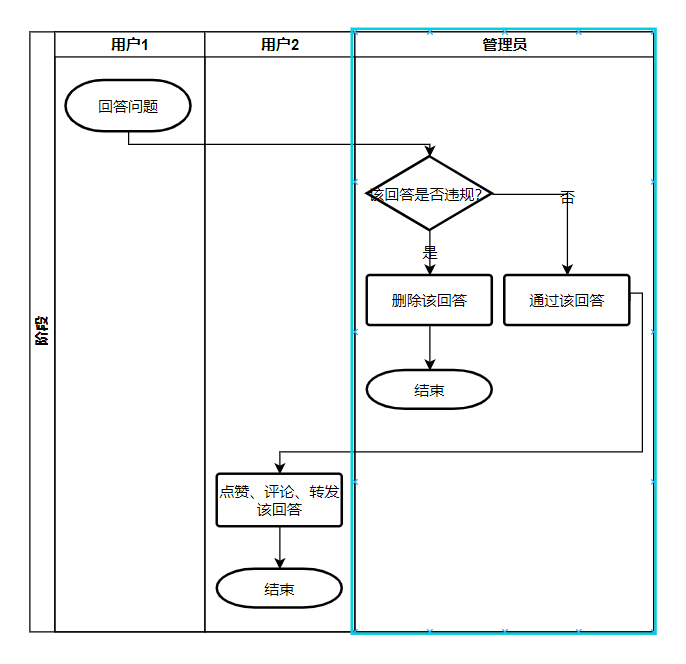
现在已经搭建了多机房，目前重要的业务都是异地多活。



# 调研分析业务场景

早期的知乎是一个聚焦专业话题的小众社区，主要面向互联网圈内人士。实行社区开放政策后，知乎转变为一个面向广大网友的知识分享平台。以问答社区为核心基础，问答内容不仅有各类专业知识，还包括了大众话题、娱乐话题。同时也推出知乎Live、私家课、值乎等多种形态的知识付费业务，用户主要是发达城市有学习需求和自我提升意识的人群。

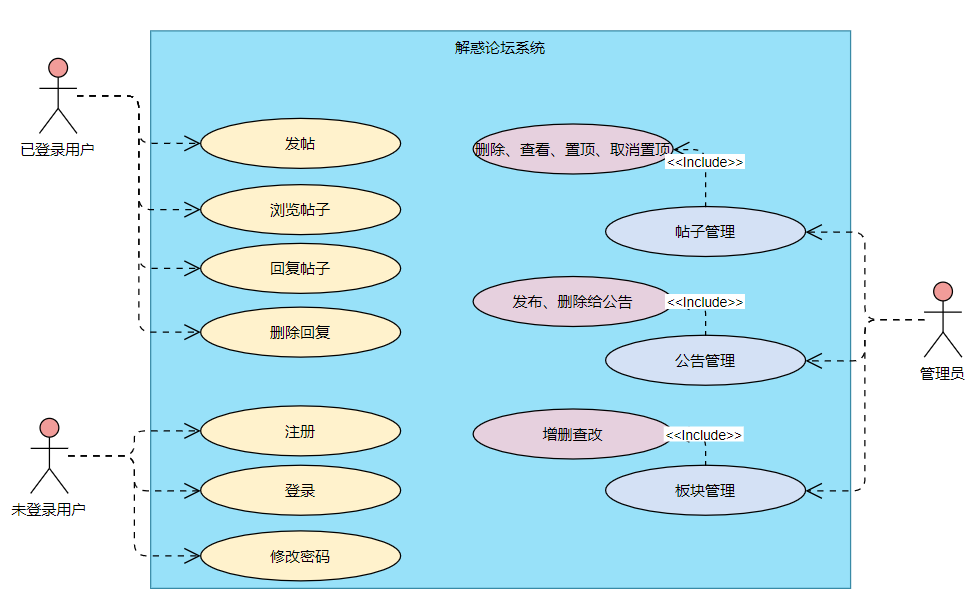
以用户发帖为例，其泳道图为：



所做的解惑论坛，包含用户界面和管理员界面。

用户界面可以让用户通过网页随时随地访问。用户登录后，可以查看今日推荐，搜索自己感兴趣的问题，并查看评论区、发表评论等。帖子可以按照其分类（全部、数据库、深度学习等）、回复数量、发布时间进行筛选。点开一个帖子后，展示该帖子的图文内容，发布时间，类别，作者及所有评论。

管理员界面大体有帖子管理、公告管理、板块管理三大功能。在帖子管理模块，可以对帖子进行删除、置顶推荐、取消置顶等；在公告管理模块，可以发布、删除公告；在板块管理模块，可以添加、查看、删除分类板块。



# 系统性能指标

1. 响应时间： 响应时间在秒级别，正常情况下1秒内能给出响应。
2. 吞吐量：1000个请求/秒
3. 资源使用率：

内存：和数据规模正相关,维持基本运行应在100MB左右。

磁盘I/O：每秒1000个请求，平均一个请求50kB，需要50Mb /秒的磁盘IO

网络I/O：同磁盘，50Mbps

cpu占用：以上述数据规模，估计需要8核\*2Ghz的CPU可以满足

1. 并发用户数：500

# 系统战略

自移动互联网以来，智能手机的出现不断的改变人们传统的生活，人们在移动端进行消费、娱乐已成常态，而除此之外，人们对知识和信息的获取也从传统的电视，书本转移到了线上的信息获取平台。知乎作为人们获取信息知识的平台，也在不断积累用户，迅猛发展。

知乎是一个获取、分享、传播知识与经验的UGC社区，“高质量”是知乎的标签。

知乎使用过的slogan：与世界分享你的知识、经验和见解。在之前的几年里，知乎相对比较注重社区内容的创造，维持了一个高标准、高质量的社区生态。过去累积的大量优质内容，使得不管是在百度搜索还是微博营销号，都能看到知乎问答内容。

2018年，知乎推出新的slogan：有问题，上知乎。无论是大众话题还是专业知识，知乎都能在最短的时间内给用户一个最全面的答案，这是普通搜索引擎所不能做到的。知乎希望培养用户“有问题，上知乎”的习惯，挖掘和培养购买知识付费服务的潜在用户。

从用户定位来看，过去知乎着重打造专业知识社区，目标用户主要是一些高知用户。而现在知乎问答社区呈现出泛娱乐化的趋势，知乎的目标用户范围正在不断扩张。

从产品定位来看，知乎不仅想做问答社区的第一，还想在知识付费领域做出一番成绩。

在线上知识服务类市场逐渐形成之后，如何变现一直是企业不断探索的问题。知乎作为这其中的佼佼者，除了传统的互联网广告的变现形式，知乎更是做了如：值乎，电子书，知乎live等多方面探索，其最近推出的知乎live更是受到广大媒体的认可，或将成为日后变现的主要形式。

# 投资和运营成本

投资和成本：前期的建设投资包含人力成本、系统开发的成本、服务器的购置成本，后期的运营成本包括：产品开发和迭代更新维护、服务器运行和维护的消耗、人力和团队管理支出、营销推广费用等等。

收益：首先可以接入传统的互联网变现形式：广告。其次开拓知识付费的市场，推出值乎、知乎live等付费栏目。

# 技术可行性

硬件环境使用X86-64通用计算机服务器，搭建集群来应对增长后的大规模用户数据和请求。在起步和小中型阶段可以与大型公有云服务商合作以加速开发，在成长得足够大规模后，必须拥有自己的云服务系统，为自己的业务专门优化来节省成本和提升用户体验。

所搭建的解惑论坛，使用基于Python语言的Django框架，以mysql为数据库，bootstrap为前端框架，以网页作为主要交互方式。现有技术和框架可以支持预想的功能实现。

# 专家论证

郑专家：

专家论证建议：经分析，该系统满足用户点单需求的基本功能，能为用户提供较好体验，在页面交互与响应速度等方面都能为用户带来良好体验，开发方面项目采用django框架可快速处理任务且安全性较高。稍有不足的是在市场定位方面竞争性较弱。

结论：该数据库应用系统的设计开发是可行的。

李专家：

专家论证建议：进行了详细的系统规划，对并发情况、系统战略等有清晰认识。

结论 ：方案具有较强可行性。