# 大型网站技术架构

## 参考教材

1、《大型网站技术架构-核心原理与案例分享》—李智慧

2、《大型分布式网站架构-设计与实践》—陈康贤

电子书地址：

## 目录

1. 大型网站架构演化
2. 大型网站架构模式
3. 大型网站核心架构要素
4. 瞬时相应：网站的高性能架构
5. 万无一失：网站的高可用架构
6. 永无止境：网站的伸缩性架构
7. 随需应变：网站的可扩展架构
8. 固若金汤：网站的安全架构
9. 淘宝网的架构演化案例分析
10. 维基百科的高性能架构设计分析
11. 海量分布式存储系统Doris的高可用架构设计分析
12. 网购秒杀系统架构设计案例分析
13. 大型网站典型故障案例分析
14. 架构师领导艺术
15. 网站架构师职场攻略
16. 漫画网站架构师

### 章节1-大型网站架构演化

现有的大型网站感官：淘宝、京东、百度、12306。

互联网的一大特点是快：业务变化快、用户增长快—>导致了996。

#### 1.1、大型网站系统特点

高并发、大流量：双11，微信红包和“有”空间红包。

高可用：7\*24小时，小问题会变成大新闻-阿里的电缆、百度域名的黑客劫持事件。

海量数据：facebook每周照片数量10亿，google百万台服务器管理数据。

用户分布广，网络情况复杂：三网合一，中美海底光缆，政治问题…。

安全环境恶劣：csdn、yahoo、查开房。

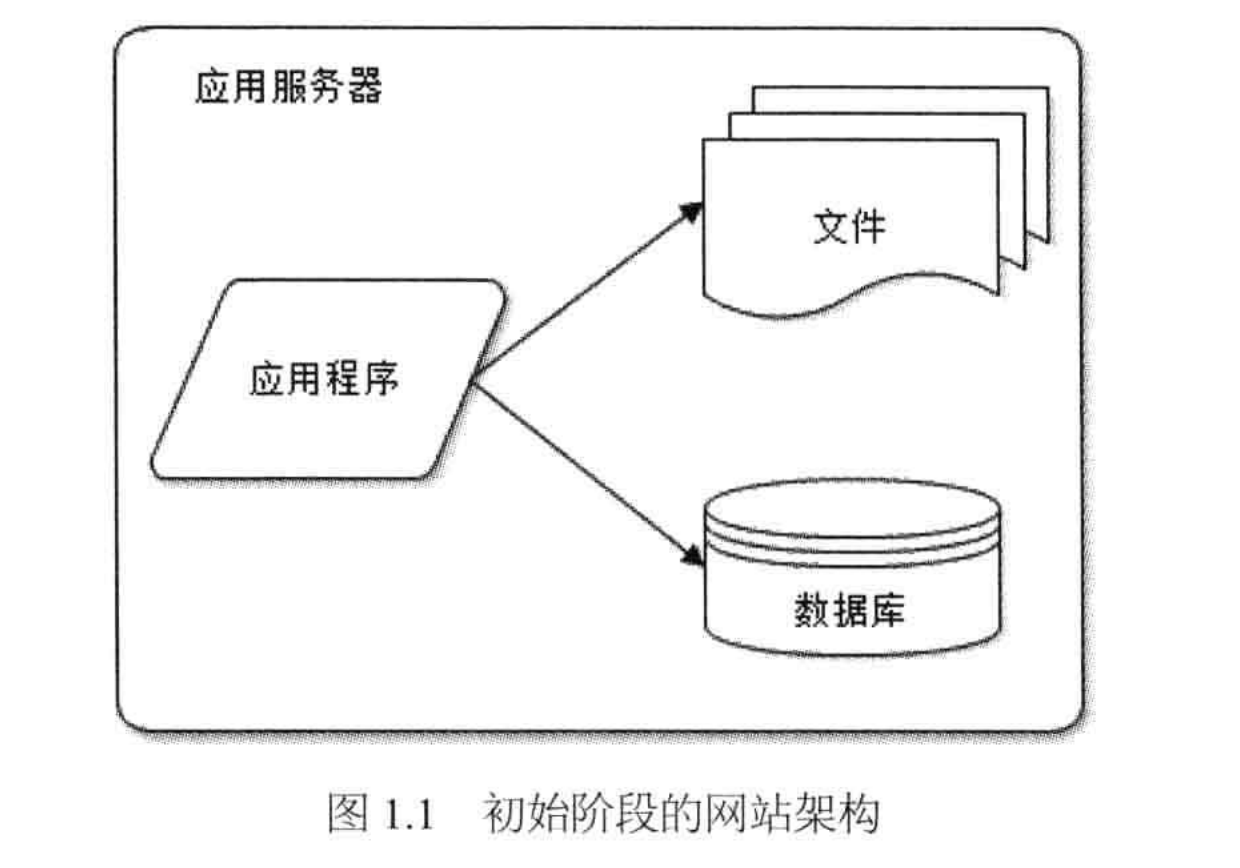
需求快速变更，发布频繁：快速响应市场，每周都有新版本，每天都有多个新特性。

渐进式发展：根据经验，好的互联网产品是逐步运营出来的，不是一步开发好。

#### 1.2、大型网站架构演化发展历程

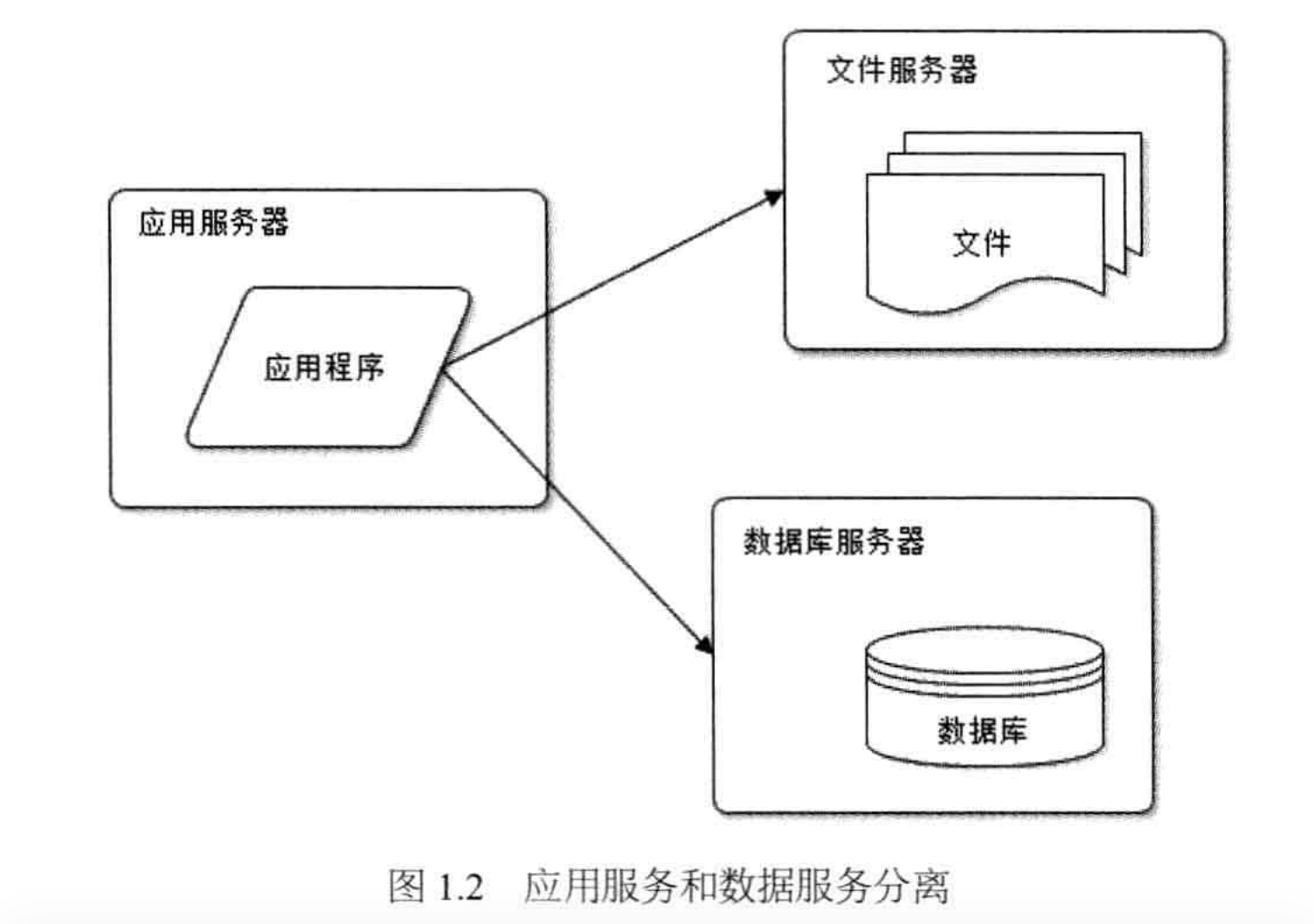
大型网站的技术挑战主要来自：庞大的用户，高并发访问和海量的数据，任何简单的业务一旦要处理数以P计的数据和面对数以亿计的用户，问题就会变得很棘手，大型网站架构主要就是解决这类问题。

1）、初始阶段的网站架构Linux+PHP+Apache+MySQL



新问题：CPU、存储空间、磁盘速度

2)、应用服务和数据服务分离



新问题：数据库压力大

3）、使用缓存改善网站性能

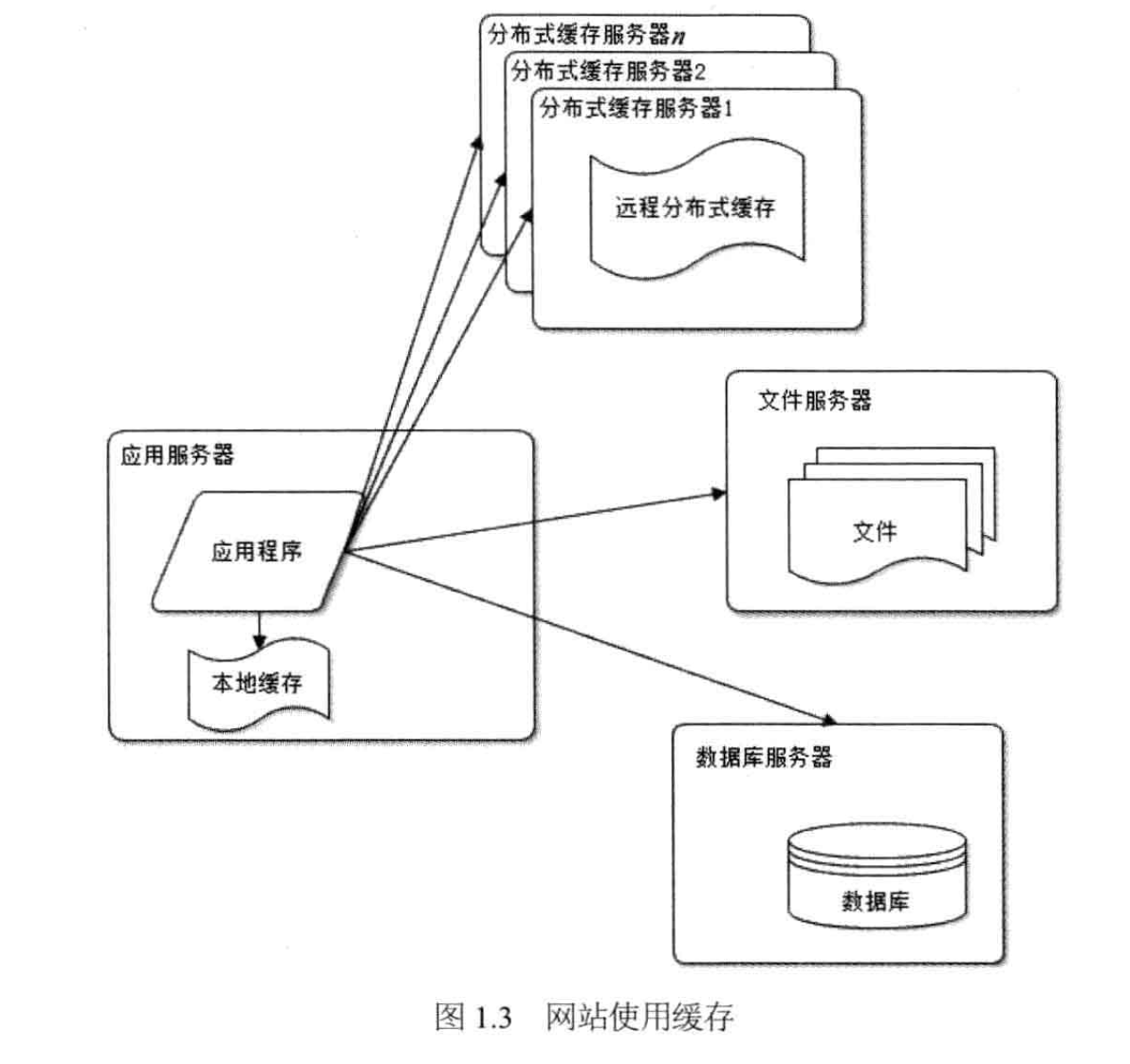
2-8法则：80%的业务访问集中在20%的数据上。淘宝的商品，百度的搜索关键词，微博账号等。

缓存类型：

应用服务器上的本地缓存和缓存在专门的分布式缓存服务器上的远程缓存。

本地缓存速度更快，但受应用服务器内存影响，数量有限。

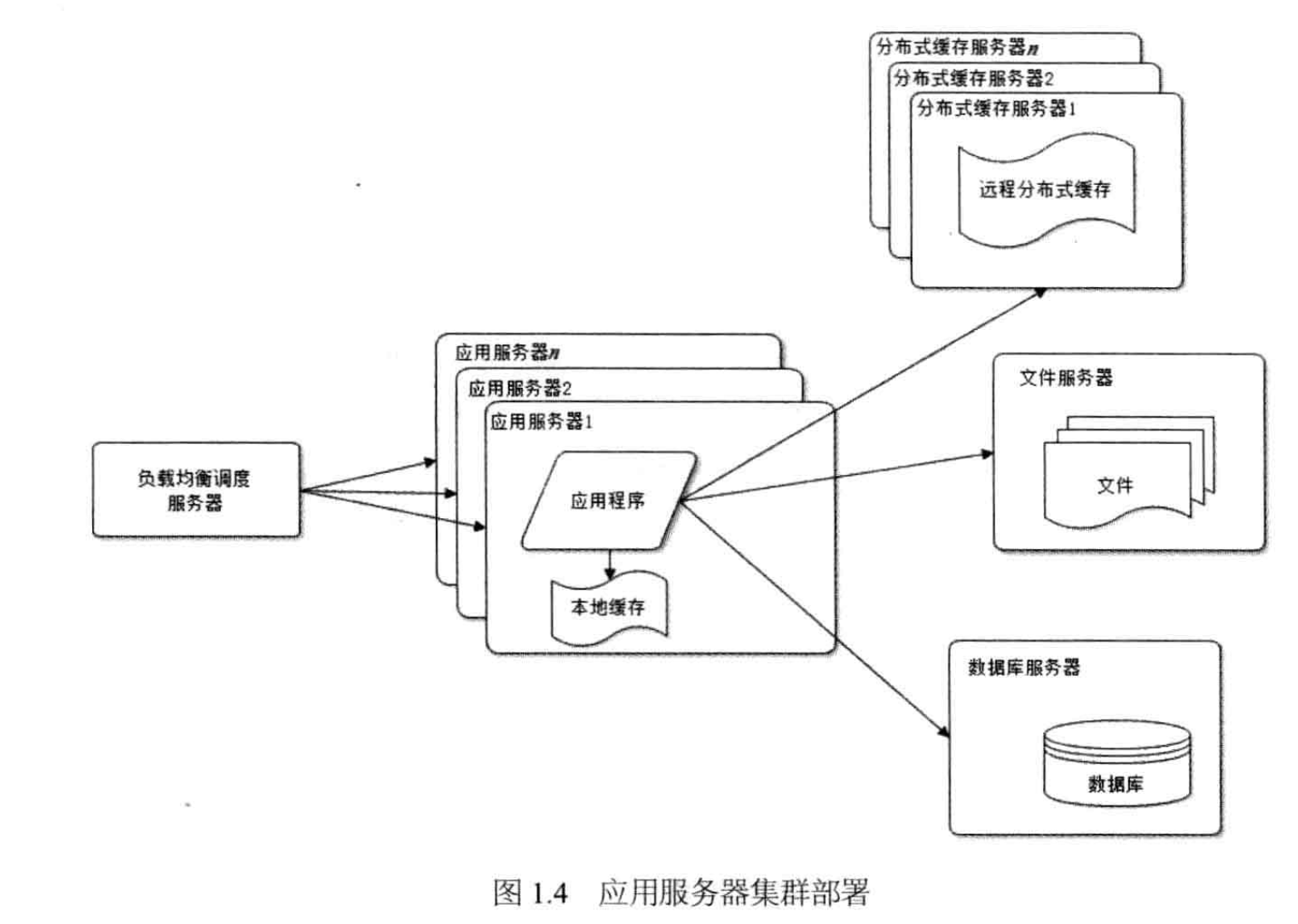
远程分布式缓存可以使用群集方式，部署大内存的服务器作为专门的缓存服务器，可以在理论上做到不受内存容量限制的缓存服务。



新问题：数据库访问压力降低，应用服务器处理请求有限。

4）、应用服务器群集改善网站的并发梳理能力

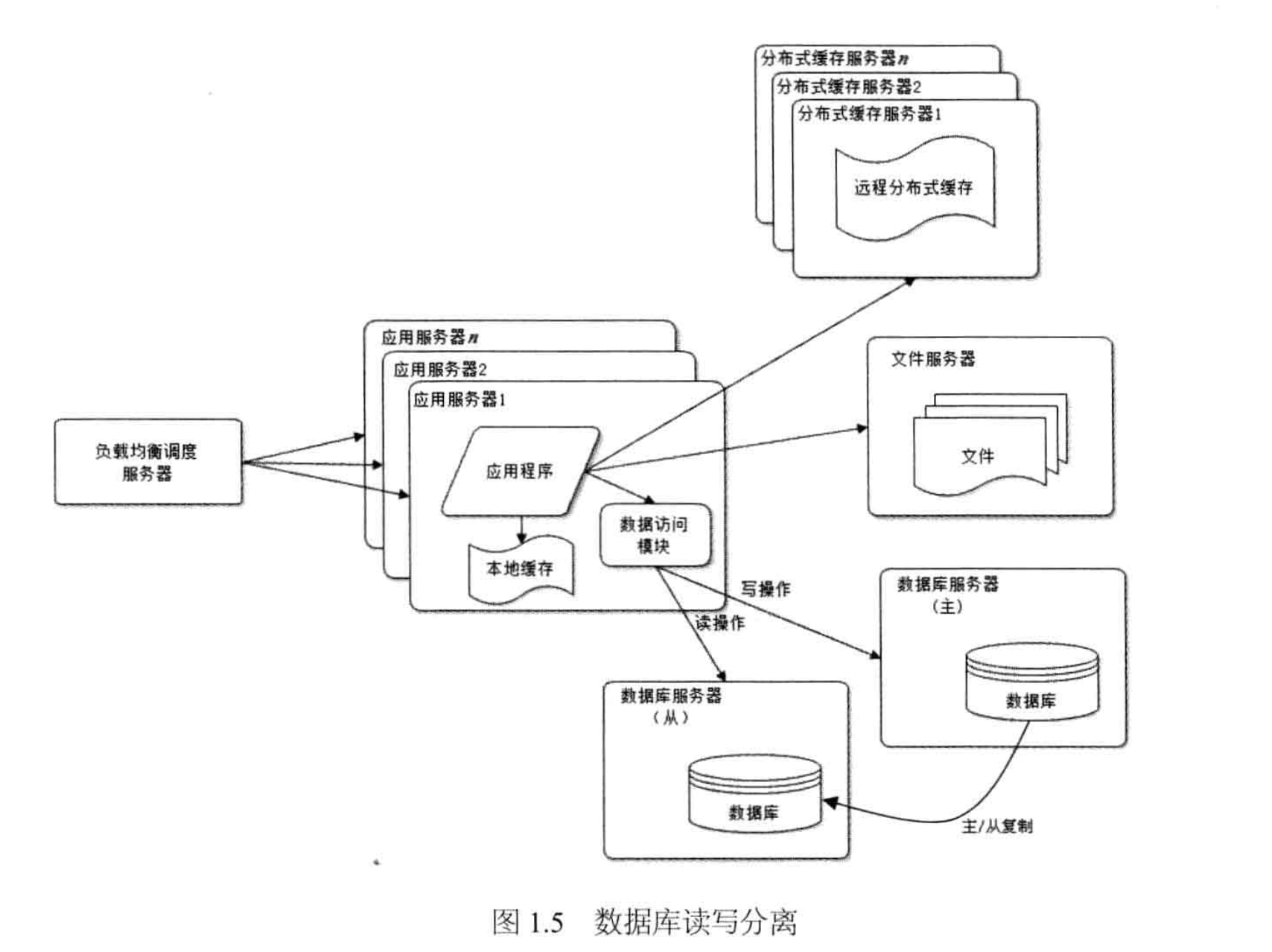
使用群集是网站解决高并发、海量数据问题的常用手段。



新问题：数据库大部分读操作通过缓存来完成，但部分读操作（缓存没有命中或过期）和所有写操作还是需要访问数据库，数据库再次成为瓶颈。

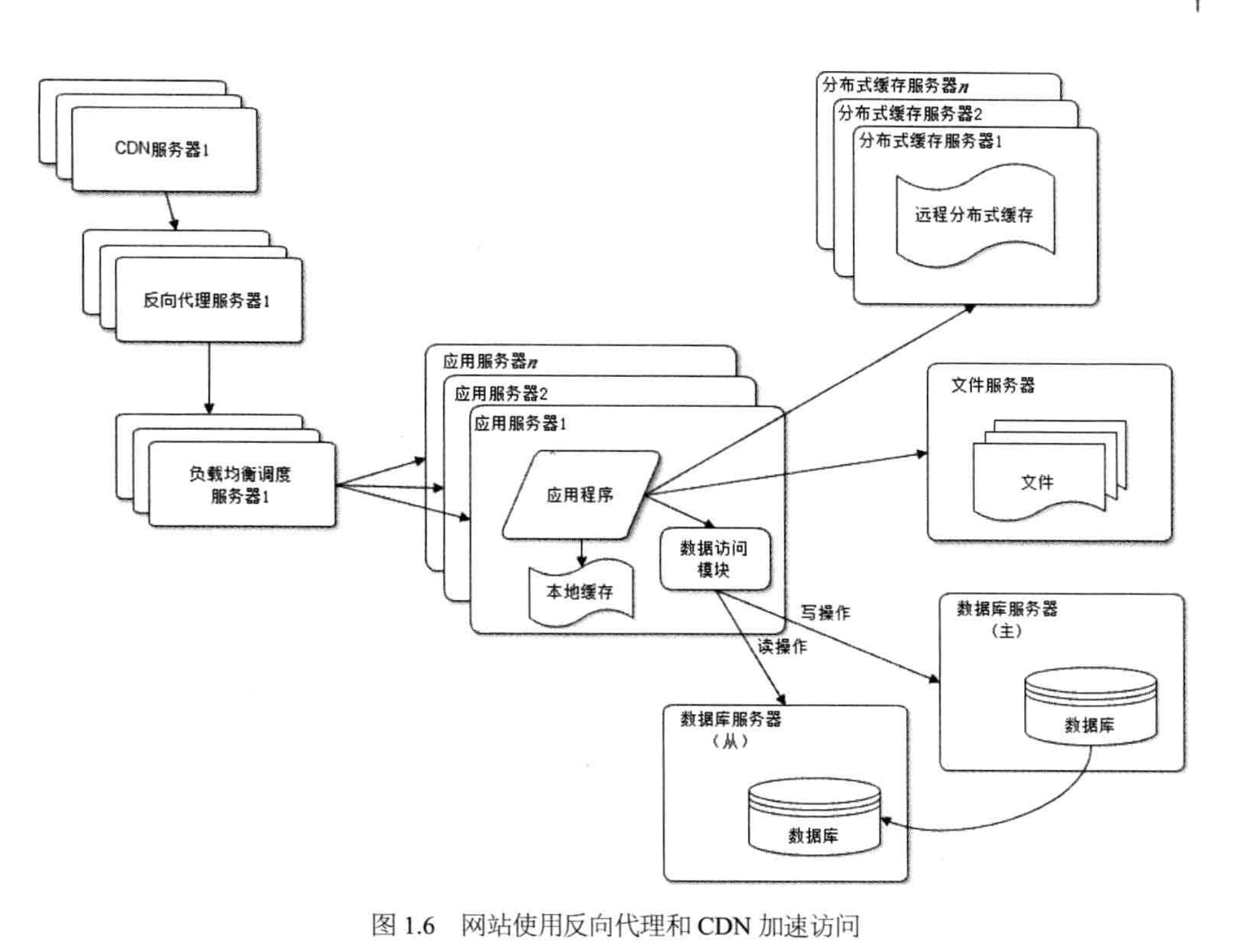
5）、数据库读写分离

数据库主从热备，实现读写分离。



6）、反向代理和CDN加速网站相应

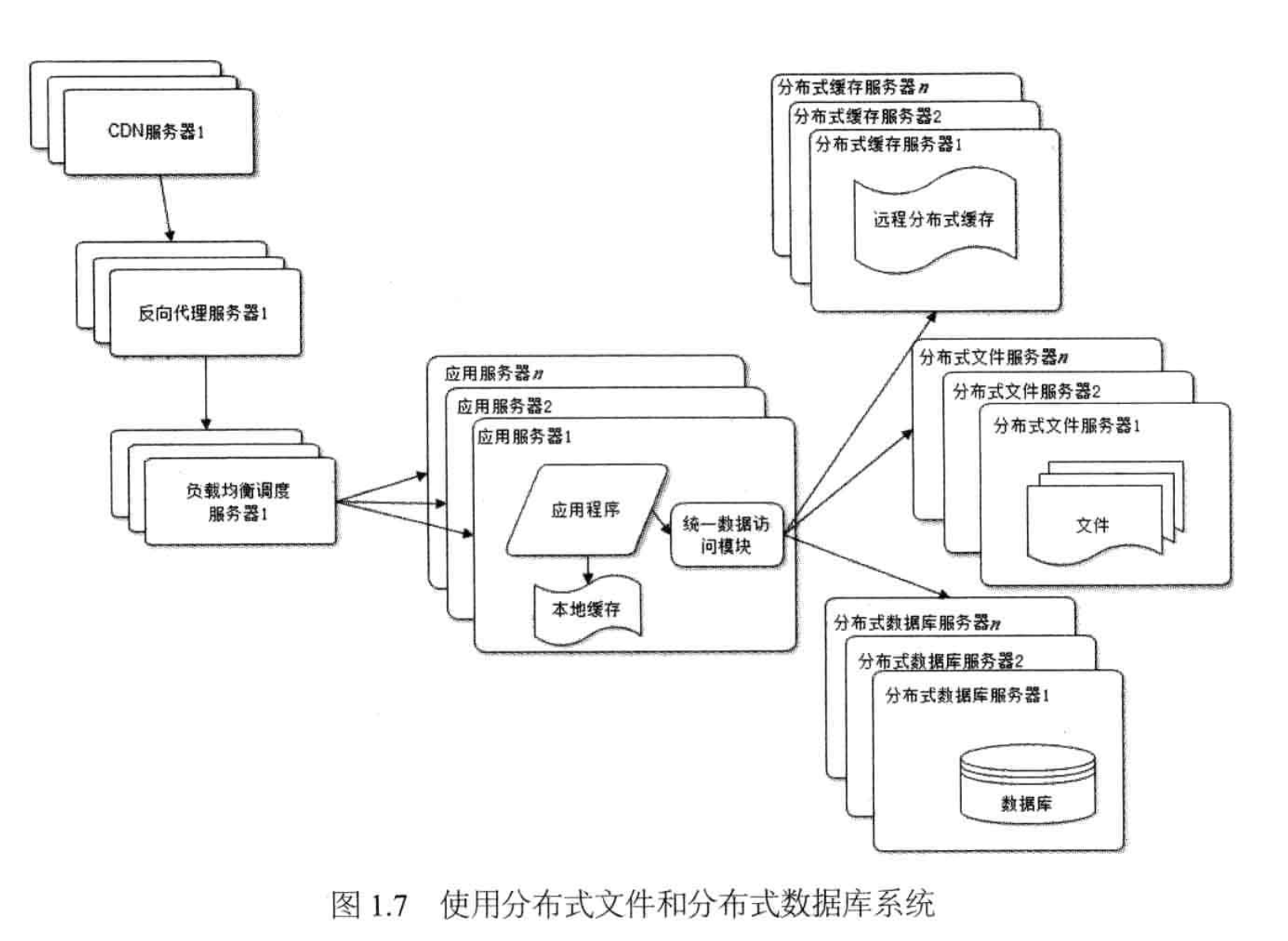
中国复杂的网络环境，不同地区用户访问网站时速度差异巨大，网站访问延迟和用户流失率有函数关系。如何加速网站访问？反向代理和CDN。



新问题：

7）、分布式文件系统和分布式数据库系统

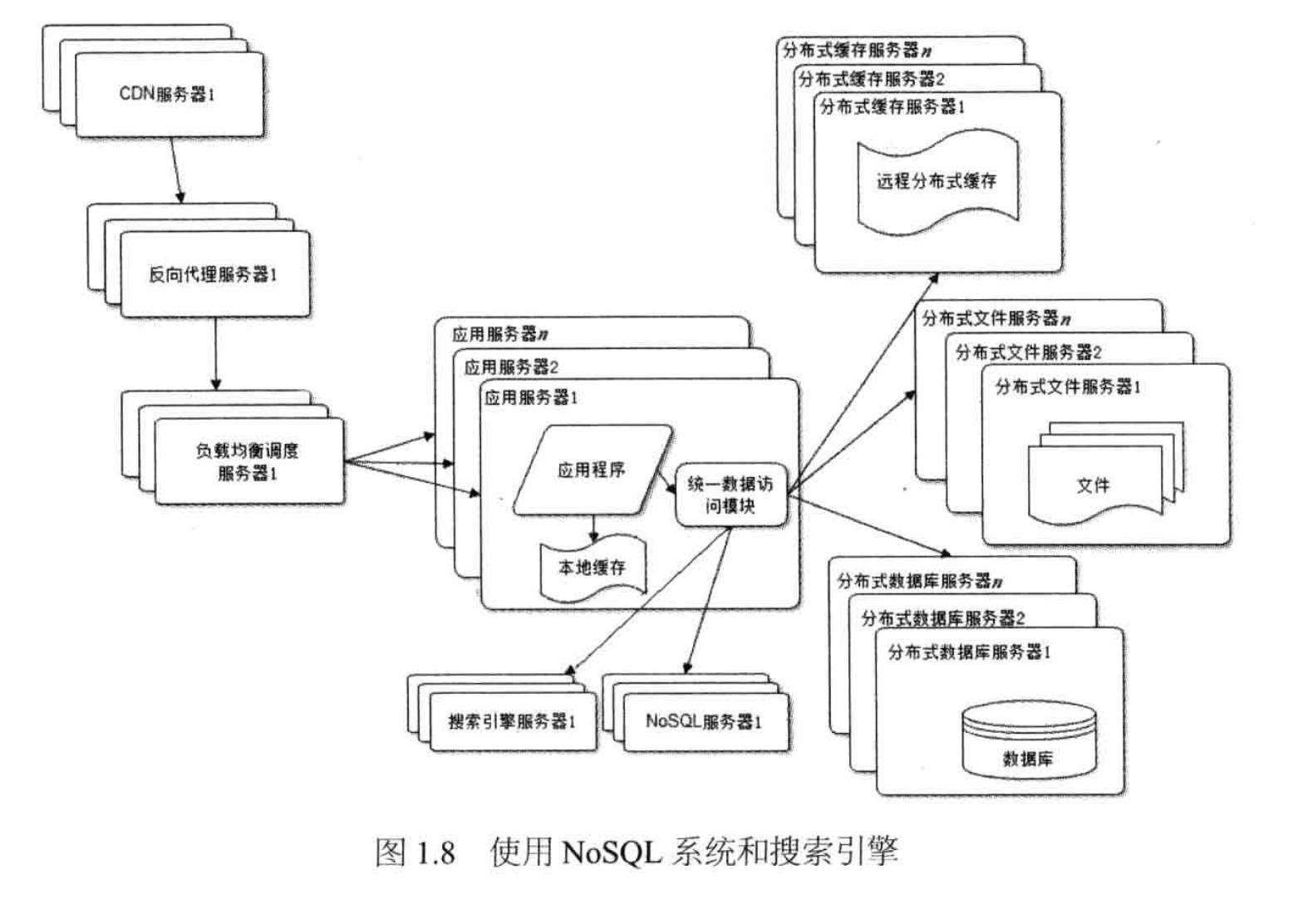
分布式数据库是数据库拆分的最后手段，常用做法是业务拆分。



新问题：

8）、NoSQL和搜索引擎

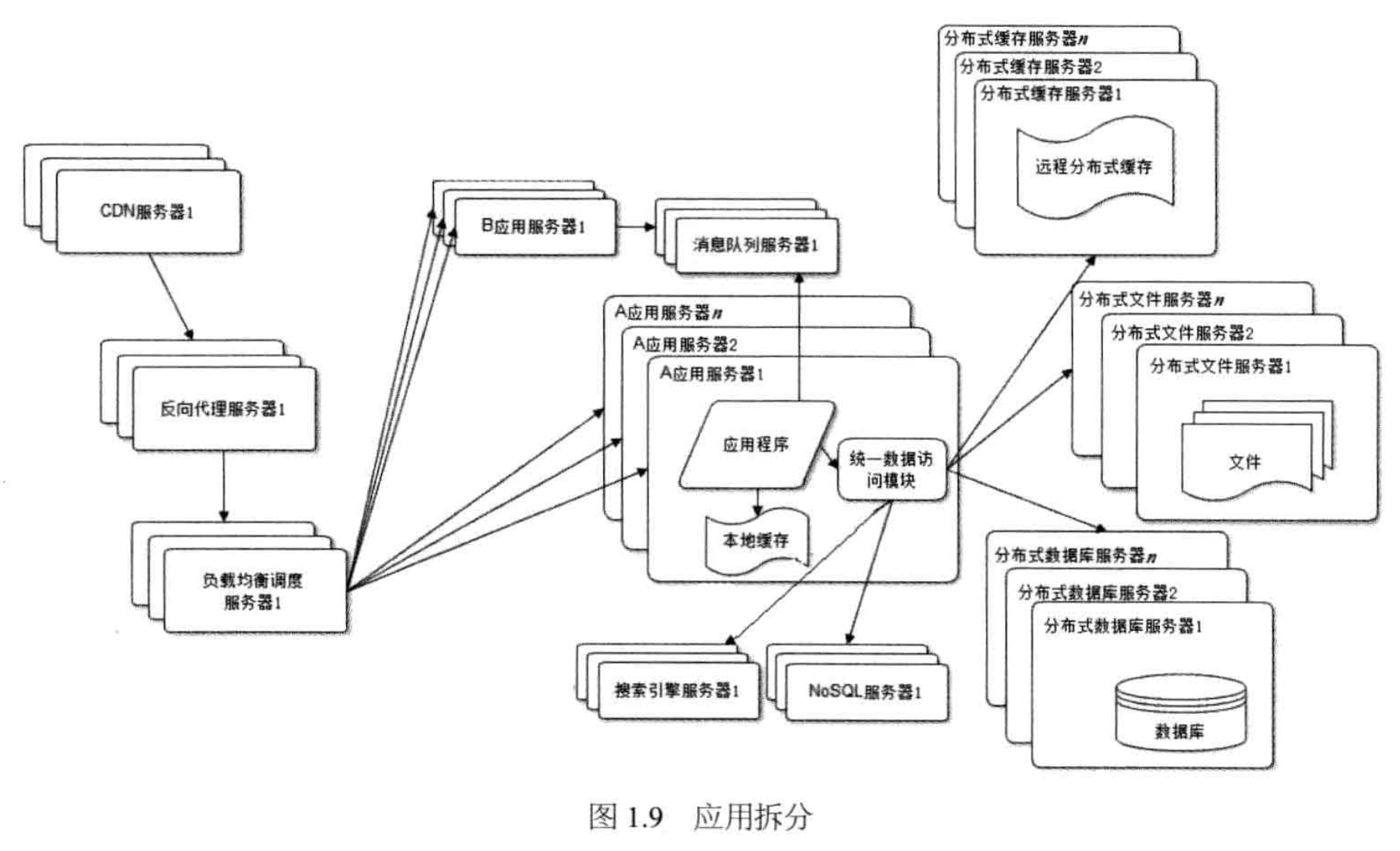
随着数据量和查询越来越复杂，关系数据库出现了瓶颈。需要非关系型数据库和搜索引擎查询方式。



NoSQL对分布式特性有更好的支持，减少应用的复杂度。

9）、业务的拆分

日益复杂的业务场景逐步将整个网站分成了不同的产品线，比如首页、商品、订单、买家、卖家、仓库、物流等等，拆分到不同的业务团队负责。技术上会根据产品线划分，拆分成不同的应用，每个应用独立部署，应用之间通过URL或消息队列建立关系，但往往是通过访问同一个数据存储来构成一个完整系统。

10）、分布式服务

业务拆分越来越小，系统存储越来越大，系统的整体复杂度呈指数级上升，部署和维护越来越困难，由于所有应用都需要和数据相连接，成千上万的应用服务器连接到数据库服务器上，数据库连接资源不足。

最终把应用和存储再次拆分，独立成微服务（原文缺陷），服务之间通过接口来访问。

#### 1.3、大型网站架构演化的价值观

1）小型网站最需要做的是为用户提供好的服务来创造价值，活下去，架构不是追求的第一目标。

2）大型网站架构技术的核心价值观是随网站所需灵活应对，是一个演化过程。

3）驱动大型网站技术发展的主要力量是网站的业务发展。

#### 1.4、网站架构设计误区

1）一味追求大公司的解决方案，借鉴学习而不是盲从。

2）为了技术而技术，脱离业务发展的实际，最求时髦的新技术。

3）企图用技术解决所有问题。12306，有空间红包。

问题：

应对2-8法则最常用的手段是什么？