



A 1st introduction to **High Performance Computing** Concepts & Techniques

Chapter 8: Case study **E-commerce system (MPMD)**



孔令波

mlinking@126.com

+86 15010255486

Outline of this course

- **Introduction**
- **Developing environments (Tools)**
- **HPC with an example**
 - Weather Forecasting
 - Python code for Sequential processing (Paradigm for Concurrent/Parallel programming)
- **Evolution of HPComputers**
- **Distributed OS?**
- **Programming frameworks**
 - MPI, CUDA, MR for Big Data, (Graph?)
- **Optional**
 - GRID,P2P(Blockchain etc.), Clouding
 - **Big e-commerce system@JD**
 - Recommending system
 - DL@GPUCUDA



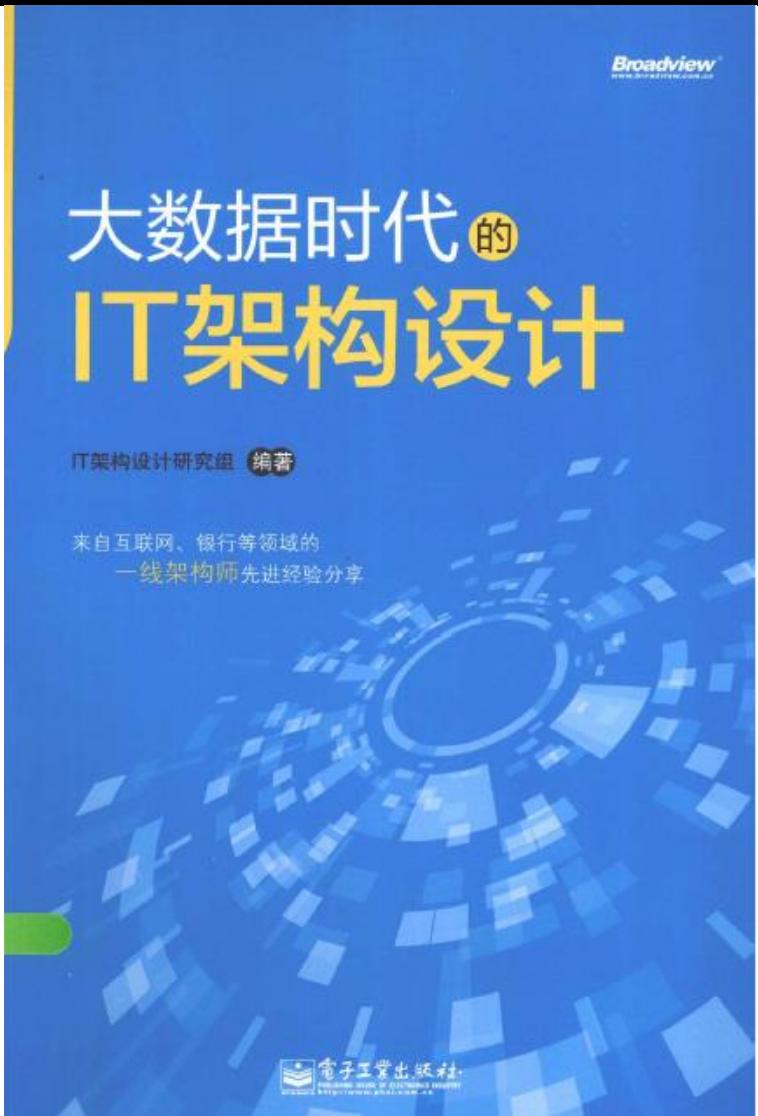
卸任50余家公司高管，刘强东江湖再见！

2020-04-14 21:01:50



- 一月前，刘强东密集卸任近20家子公司的CEO之后，最终还是走下了京东集团CEO的高位。
- 有人说，刘强东仍然以78%的股权掌握着京东的权柄；也有人说，在刘强东这个年纪，马云、俞敏洪、马化腾还在一线拼着。相比于马云、柳传志的“退休”，刘强东的隐退的确暗淡了许多。
- 《圆月弯刀》的结局里，丁鹏在和任天行决战前，把寄托心安的无命客栈交到了荆无命手里。刀光剑影间，丁鹏还是消失在江湖之上，只剩传说。

大数据时代的 IT 架构设计 / IT 架构设计研究组编著. —北京：电子工业出版社，
2014.4
ISBN 978-7-121-22605-2



第 1 篇 Hadoop 平台应用架构

Hadoop 技术在电信运营商上网日志处理中的应用架构 / 方建国 ······

Hadoop 平台在金融银行业的应用架构 / 薛强彦 ······

优酷土豆 Hadoop 平台开放之路 / 傅杰 ······

第 2 篇 推荐系统架构

腾讯个性化精准推荐平台架构 / 洪坤乾 ······

社交网络大数据挖掘与社会化推荐引擎的构建 / 邓雄 ······

第 3 篇 应用架构实践

微博用户兴趣建模系统架构 / 张俊林 ······

针对移动互联网应用的网络建设和优化 / 李永斌 ······

西门子系统架构管理实践 / 李伟 ······

第 4 篇 大数据下的存储实践

如何构建高性能、稳定的 SOA 应用 / 汪洋 ······ 淘宝海量文件存储实践 / 张友东 ······

基于开源技术的网络安全架构实践 / 李晨光 ······ 内存作为统一存储实践 / 尹伟铭 ······

TOGAF 应用架构设计落地之道 / 温昱 ······

分布式系统安全设计解决方案 / 汪洋 ······

第 5 篇 数据库架构实践

数据库优化精髓暨数据架构最佳实践 / 侯圣文 ······

Oracle、MySQL 还是 NoSQL / 简朝阳 ······

SQL Server 中的高可用性 / 宋云剑 ······

第 6 篇 大数据下的 IT 架构变迁

京东商城网站架构演变 / 王晓钟 ······

庶民架构——新浪视频后台架构变迁 / 郝光普 ······



从本书中你将收获到的不仅仅只是架构理论。
更多的来源于互联网技术下大规模架构和演进过程中
核心技术实践的解决方案。

Broadview
WORLD PUBLISHING COMPANY

人人都是架构师

分布式系统架构落地与瓶颈突破

高阳龙 著



中国工信出版集团



电子工业出版社

www.java1234.com



大型网站技术架构

核心原理与案例分析

李智慧 著

电子工业出版社
Publishing House of Electronics Industry
北京·BEIJING

www.java1234.com

仅供非商业用途或交流学习使用

Broadview®
www.broadview.com.cn



大型网站 技术架构演进与性能优化

许令波◎著



中国工信出版集团



电子工业出版社

www.java1234.com





- 互联网运营智慧
- 作者: 田逸
- 出版社: 无 9787302239697
- 副标题: 高可用可扩展网站
技术实战
- 出版年: 2011-1
- 页数: 397
- 定价: 69.00元
- ISBN: 9787302239697



□ 高性能响应式Web开发实战



□ 高性能电子商务平台构建:架 构、设计与开发

Chapter 8: Case Study

□ **History of JD**

□ **Overview of used techniques in JD**

□ **Architecture**

□ **Software architect [软件架构师]**



<https://www.tianyancha.com/>

https://www.tianyancha.com/company/12562796

345网址导航 - 开... Google 翻译 盘优搜,百度网盘搜... 西林街搜索 (X 站... 简书 - 创作你的创作 天眼查-商业安全工... 书格 OK电影院 https://ambry.pw Ebook-dl | Free E

天眼查 都在用的商业安全工具
国家中小企业发展基金旗下

查公司 查老板 查关系
北京京东世纪贸易有限公司 天眼一下

VIP会员 NEW 百宝箱 合作通道 登录/注册

官方信息 自主信息 18

北京京东世纪贸易有限公司 我要认证

在业 中概股 | 京东 JD

电话: 8911**** 查看更多 编辑 邮箱: ***@126.com 查看更多 认证成功后即可编辑信息, 去认证

网址: www.jd.com 地址: 北京市北京经济技术开发区科创十一街18号C座2层201室 附近公司

简介: 北京京东世纪贸易有限公司是一家网络零售企业。

产品信息: 京致衣橱 融资历程 (0) 竞品数量 (50)

更新时间: 2019-03-29

发票抬头 下载报告 我要投诉 收藏

天眼风险 自身风险 783 周边风险 999+ 预警提醒 208 查看详情 **情报动态** 2019-03-28 新增 司法拍卖

公司背景 296 司法风险 999+ 经营风险 495 公司发展 424 经营状况 104 知识产权 999+ 历史信息 265 VIP 问大家



都在用的商业安全工具
国家中小企业发展基金旗下

查公司

查老板 查关系

北京京东世纪贸易有限公司

天眼一下

VIP

官方信息

自主信息 18

公司背景 296

司法风险 999+

经营风险 495

公司发展 424

经营状况 104

知识产权 999+

历史信息 265 VIP



刘强东

他有 409 家公司，分布如下

江苏 (共112家) 宿迁大荣恒元管理顾问...等
北京 (共81家) 北京京东振世信息技术...等
其他 (共216家) 江苏京东信息技术有限...等

股东

高管

XXXX有限公司

对外投资

分支机构

历史股东

历史法人

查看详情

登录查看

北京京东世纪贸易有限公司

京东香港国际有限公司
股比: 100% 认缴金额: 1397985564万美元

注册资本 ②

139798.556万美元

成立日期 ②

2007-04-20

评分 80



经营状态

在业

工商注册号 ②

110000450007248

统一社会信用代码

911103026605015136

组织机构代码 ②

660501513

纳税人识别号

911103026605015136

公司类型

有限责任公司(台港澳法人独资)

营业期限

2007-04-20至2037-04-19

行业 ②

批发业

纳税资质

-

核准日期

2017-11-30

实缴资本

-

人员规模

5000-9999人

参保人数

8178

登记机关 ②

北京市工商行政管理局北京经济技术开发区分局

注册地址 ②

北京市北京经济技术开发区科创十一街18号C座2层201室 附近公司

英文名称

Beijing Jingdong Century Trading Co.,Ltd.



1998年6月18日 刘强东在中关村创立“京东公司”代理光磁产品

2001-2003年，最多时北京、上海、沈阳12家连锁店，营业额6000万元/年。

2001年，成为中国最大的光磁代理商，第一家零售店“京东多媒体”在中关村苏州街上的银丰大厦开张。

2003年非典，放弃了要在全国扩张门店的计划并开始尝试在网上发帖售卖商品。

2004年1月京东涉足电子商务领域，京东多媒体网正式开通，启用域名 www.jdlaser.com。

2005年10月京东坚决地关闭了所有门店，实行100%的线上销售。

2005年11月京东多媒体网日订单处理量稳定突破500个。

2005年关闭所有线下零售店 3000万元/年

2006年1月京东宣布进军上海，成立上海全资子公司。

2007年5月京东广州全资子公司成立，全力开拓华南市场。广州全资子公司的成立代表着京东由北京、上

海、广州三地为基础覆盖全国的销售网络的形成。



2007 年 6 月京东正式启动全新域名 www.360buy.com，并成功改版。日订单处理量突破 3000 个。

2007 年 7 月京东建成北京、上海、广州三大物流体系，总物流面积超过 5 万平方米。

2007 年 8 月京东赢得国际著名风险投资基金——今日资本的青睐，首批融资千万美金。

2007 年 10 月京东商城在北京、上海、广州三地启用移动 POS 上门刷卡服务，开创了中国电子商务的先河。

2008 年 6 月京东商城在 2008 年初涉足销售平板电视，并于 6 月将空调、冰洗、电视等大家电产品线逐一

扩充完毕。标志着京东公司在建司十周年之际完成了 3C 产品的全线搭建，成为名副其实的 3C 网购平台。

2008 年 7 月京东商城作为国内最大的 3C 网购平台，凭借多年来惊人的高速发展首次参选“清科-2008 中国最具投资价值企业 50 强”便登上了榜单。

2009 年 01 月京东商城获得来自今日资本、雄牛资本以及亚洲投资银行家、“红筹之父”梁伯韬的私人公司共计 2100 万美元的联合注资。

2009 年 3 月，斥资 2000 万元自行成立了上海圆迈快递公司，今年拟投资 6700 万元自建物流公司。



2009 年第一季度，易观国际发布的《2009 年第一季度中国 B2C 网上零售市场季度监测》数据显示，3C 产
品 B2C 网上零售市场销售规模已占 B2C 整体市场 38.19%，排名商品品类销售规模市场榜首，其中京东商城
以 5.7 亿元销售额位列 B2C 市场第一。

2009 年 6 月，销售额达到 3.78 亿元，同比增长 270%，平均订单金额达到了 800 元。2010 年 4 月，京东
商城在北京等主要城市推出“211 限时达”配送服务，在全国实现“售后 100 分”服务承诺，随后又推出
“全国上门取件”、“先行赔付”、7×24 小时客服电话等专业服务。京东商城的服务系统正在逐步实现
跨越性的升级。

2010 年 5 月，公司宣布将在北京新建一个占地 30 万平方米的物流中心，成为亚洲最大电子商务物流中心，
可满足日处理 10 万订单，年销售额超 200 亿的需要。

2010 年底，京东商城第三方销售平台“品牌直销”频道正式上线，目前约有 500 个品牌，近 10 万种商品
入驻京东商城。平台商户可以分享京东的仓储、配送、客服、售后、货到付款、退换货和自提货等服务，
进一步减少自建服务体系的成本。

2010 年底以 20 万收购了域名 TopLife.com，作为京东的奢侈品垂直 B2C 网站。
2010 年，京东商城营业额 102 亿元，亏损率为 8%。京东商城预计 2011 年公司的销售额将翻倍增长，但
收入构成将呈现多元化。预计营业额将达到 240—260 亿元，其中 IT 产品 85 亿元，手机和数码类产品 60
亿元，大家电 50 亿元、小家电 25 亿元、日用百货 35—40 亿元。



-
- 2010年12月23日，京东商城团购频道于12月23日正式上线，京东商城注册用户均可直接参与团购。
 - 2011年4月1日，刘强东宣布完成C2轮融资，投资方俄罗斯的DST、老虎基金等六家基金和一些社会知名人士融资金额总计15亿美元，其中11亿美元已经到账。
 - 2012年5月29日，京东商城集团旗下日韩品牌综合类网上购物商城——迷你挑正式上线！
 - 2012年11月27日，京东叫板阿里巴巴 涉水供应链金融领域。
 - 2013年3月30日19点整正式切换了域名，随后京东商城发布新的logo和吉祥物。

- 在2013年7月29日，京东POP开放平台大会上，京东集团财务副总裁范微表示，未来针对POP平台上的卖家，京东将提供小额贷款、流水贷款、联保贷款、票据兑现、应收账款融资、境内外保理业务等金融服务。
 - Platform Open Plan，直译成中文是“平台开放计划”，也就是京东原来是做自营（自己开店买自己家的东西），后来将平台开放给第三方商家。
- 2013年12月26日工业和信息化部公布批准通过2013年移动通信牌照审核企业名单，京东集团正式获得虚拟运营商牌照，成为国内获得批文的企业。京东集团作为国内的综合网络零售商，将结合自身特点，为用户提供个性化的移动通信业务。 [13]

- 2014年1月30日，京东向美国证券交易委员会（SEC）承报了拟上市的F-1登记表格，美银美林和瑞银证券为主要承销商。^[16] 2月19日，腾讯在港交所发布公告，称入股大众点评，占股20%，但没有公布具体入股金额。在腾讯入股后，大众点评的商户信息、消费点评、消费优惠、团购、餐厅在线预订、餐饮外卖等将与QQ和微信等产品深度结合。^[17] 3月10日，腾讯港交所公告，同意约2.15亿美元收购京东3.5亿多股普通股股份，占上市前在外流通京东普通股的15%。同时京东腾讯还签署了电商总体战略合作协议，腾讯旗下拍拍C2C、QQ网购等附属关联公司注册资本、资产、业务转移予京东。^[18]

- 2014年4月2日，京东集团正式进行分拆，其中包括两个子集团、一个子公司和一个事业部，涉及金融、拍拍及海外业务。^[19] 具体的分拆方式是：京东集团旗下设京东商城集团、金融集团、子公司拍拍网和海外事业部，京东创始人刘强东会担任京东集团CEO。^[19]
- 2014年5月22日上午9点，京东集团在美国纳斯达克挂牌上市（股票代码：**JD**）。**美国也迎来了中国最大的赴美IPO。** 京东开盘价21.75，较发行价上涨14.5%，并且开盘之后一路上涨，截止2014年5月，京东市值超过300亿美元，且在中概股中排名第二。^[20] 京东董事局主席刘强东敲响上市钟，发行价19美元，按此计算，京东市值为260亿美元，成为仅次于腾讯、百度的中国第三大互联网上市公司。京东商城登陆纳斯达克首日，开盘价21.75美元，较19美元的发行价上涨14.5%，报收于20.90美元，较发行价上涨10%。^[15] ^[21]

- 2014年10月20日，京东宣布其位于上海的首个“亚洲一号”现代化物流中心（一期）在双十一大促前夕正式投入使用，京东物流战略中又一重点举措落地。 [22]
- 2014年11月，京东集团宣布大家电“京东帮服务店”在河北省赵县正式开业。京东称，未来3年，“京东帮服务店”将在全国区县铺开，达到千余家。京东表示，此举主要是帮京东渠道下沉，藉此，京东大家电可在四~六线城市进行物流提速。 [23]
- 2014年，京东市场交易额达到2602亿元，净收入达到1150亿元。 [25]

- 2015年8月7日，京东发布财报，宣布董事会2015年5月批准针对公司董事长兼CEO刘强东的一项为期10年的薪酬计划。根据该计划，**刘强东在计划规定的10年内，每年基本工资为1元，且没有现金奖励。**同时宣布，针对原采销体系组织架构进行了事业部制调整，调整后分为四大事业部，分别是3C、家电、消费品和服饰家居，分别由京东集团副总裁王笑松、闫小兵、冯轶、辛利军担任事业部总裁，直接向京东商城首席执行官沈皓瑜汇报。
- 此外，京东还宣布与永辉超市建立战略伙伴关系，在采购、O2O、金融、信息技术等方面拟构建互为优先、互惠共赢的战略合作模式，实现共同发展、合作共赢。 [27]

- 2015年10月17日，腾讯集团与京东集团在京联合宣布推出全新战略合作项目 - 京腾计划，双方以各自资源和产品共同打造名为“品商”的创新模式生意平台。腾讯公司CEO马化腾说，全球范围内也没有社交和购物联合的先例，这是中国两大类平台的首次尝试。 [29]
- 2015年10月10日，京东集团发布公告称，因C2C（个人对消费者）模式当前监管难度较大，无法杜绝假冒伪劣商品，决定到12月31日时停止提供其C2C模式（拍拍网）的电子商务平台服务，并在三个月的过渡期后将其彻底关闭。 [30]

- 2016年4月15日，京东集团宣布，旗下O2O子公司“京东到家”与众包物流平台“达达”合并一事达成最终协议。^[31] 京东将以京东到家的业务、京东集团的业务资源以及两亿美元现金换取新公司约47.4%的股份并成为单一最大股东。^[32]
- 2016年6月8日，《2016年BrandZ全球最具价值品牌百强榜》公布，京东首次进入百强榜，品牌价值同比增长37%至105亿美元，排名第99。^[33]
- 2016年11月23日，京东集团推出“京东物流”全新品牌标识，并正式宣布京东物流将以品牌化运营方式全面对社会开放。同时京东物流还公布了全面迈向“开放化、智能化”的战略规划，并希望借此成为中国整个商业社会的基础设施提供商。^[36]

- 2016年6月8日，《2016年BrandZ全球最具价值品牌百强榜》公布，京东首次进入百强榜，排名第99。 [33]
- 2016年第四季度，公司净营收为人民币803亿元（约合116亿美元）。 [38]
- 2018年4月，京东投资5.37亿元人民币入股安联财险中国。 [44]
- 2018年04月27日，爱奇艺与京东集团正式达成在线娱乐与电商网站会员权益互通的独家战略合作。合作达成后，用户成为爱奇艺或京东任一平台年卡会员，只需完成权益领取并激活即可同时获得相应的合作平台会员权益。即开通爱奇艺年卡会员可获得京东PLUS尊享价格、运费券礼包、专属客服、上门退换货等专享权益；开通京东PLUS会员可获得VIP独家内容、会员专属视听体验、专属活动参与、影视周边衍生特权等爱奇艺VIP会员专属权益。
[45]

- 2018年6月18日，谷歌5.5亿美元入股京东，双方将展开战略合作；同日，京东第一架重型无人机正式下线。 [47-48]
- 2018年7月12日，京东金融宣布近期已与中金资本、中银投资、中信建投和中信资本等投资人签署了具有约束力增资协议计划，融资金额约为130亿元人民币，投后估值约1330亿人民币。 [49]
- 2018年7月24日，银保监会批复同意了安联财产保险（中国）有限公司的增资方案，注册资本从8.05亿元人民币增至16.1亿元。其中，由北京京东叁佰陆拾度电子商务有限公司认购新增注册资本中的4.83亿元。 [8]

- 2018年8月13日，京东与英特尔对外宣布深化战略合作，双方将在目前的合作基础上，正式启动全量企业级业务合作。 [52]
- 2018年8月，刘强东参加清华大学和明尼苏达联合举办的DBA项目，他和家人一起乘坐私人飞机前往明尼苏达阿伯里斯。按照计划，他要在明尼苏达参加为期一周的暑期课程。在这里，他遇到了涉事的刘姓女生。8月30日，刘强东的一位朋友，也是该DBA项目的同学，邀请刘姓女生参加一个私人晚宴，晚宴结束后，刘强东和女生一起回到其所在的宿舍，然后两人发生性关系。女生指控刘强东强暴了她，并于第二天报了警。
 - <https://tech.163.com/19/0418/08/ED1J1DNU00097U7R.html>
- 2018年9月28日，京东集团与泰国尚泰集团一起打造的泰国线上零售平台JD CENTRAL正式上线运营。 [54]

- 2019年快递“春运模式”收官，京东连续七年“春节也送货”^[59]
- 2019年，京东27亿收购北京翠宫饭店100%股权。翠宫饭店成立于1987年，注册资本4.8亿元。北京翠宫饭店于2018年11月26日在北京产权交易所挂牌出售已为其二度挂牌。挂牌转让方为海淀国资经管中心，转让底价为26.83亿元^[60]。
- 2019年2月25日，金正大集团与京东集团在京东总部大楼签署战略合作协议，深入推进“互联网+农业”。^[61]
- 2019年3月28日，京东发布iSRM智能采购管理平台，该平台针对工业制造业的长尾商品多、供应商数量庞大、供应链管理难度大等问题提出了相应的解决方案^[62]。

□ 2019年4月

- 裁员
- 996
- 混日子的不是兄弟
- 18日刘强东性侵案起诉书全文曝光 被指控6项“罪名”



□ 京东商城

- 2013年5月6日，京东商城在完成内测后，正式与消费者见面，用户可在京东上购买食品饮料、调味品等日用品。

□ 京东金融

- 2015年4月28日，京东金融宣布网银钱包更名为京东钱包，网银+更名为京东支付，京东金融还提出围绕京东支付体系，为用户提供全方位金融解决方案。
- 京东金融主要有2大拳头产品，分别为京东众筹、京东白条。

□ 京东云

- 依托京东商城电商优势而开发的京东电商云平台，正在基于其产业链优势构建一个庞大的电商云生态系统，将应用推进云计算落地的真谛演绎得淋漓尽致。

□ JIMI机器人

- JIMI(JD Instant Messaging Intelligence)是京东自主研发的人工智能系统，它通过自然语言处理、深度神经网络、机器学习、用户画像、自然语言处理等技术，能够完成全天候、无限量的用户服务，涵盖售前咨询、售后服务等电子商务的各个环节，堪称京东用户的购物伴侣。 [63]

□ 京东房产

- 2018年10月22日，我爱我家与京东房产正式签署战略合作协议。双方将充分发挥各自的资源及渠道优势，着力推进二手房、租房、长租公寓等数据共享，致力解决房产交易中信息不实、流程繁杂等行业难题。京

□ 京东农牧

- 在2018年11月20日召开的2018京东数字科技全球探索者大会上，京东宣布京东金融品牌升级为京东数字科技，旗下子品牌京东农牧正式亮相，并成立京东农业研究院，将利用人工智能技术推动养猪业升级。 [65]
- 京东农牧还通过与中国农业大学、中国农科院等机构合作，自主研发并推出京东智能养殖解决方案，并联合中国农业大学建设丰宁智能猪场示范点。 [65]

线下销售

□ 京东之家

- 2016年11月10日，国内首家“**京东之家**”在长沙平和堂开业。京东之家是京东线下服务的载体和体验店^[66]。
- 京东之家主要布局一二线城市的核心商圈，主营手机、数码、电脑等3C产品，已经在北京、上海、深圳等地开设^[67]。

□ 京东专卖店

- 京东专卖店主要设在各级城市的次商圈，京东将在全国开设超过300家以3C为主的零售体验店^[67]。

□ 京东便利店

- **京东便利店**全面接入京东支付，并正式启用全球首个低造价、可模块化组装的智能门店解决方案D-MART^[68]。

Chapter 8: Case Study

□ History of JD

□ Overview of used techniques in JD

□ Architecture

□ Software architect [软件架构师]

京东人自己的提炼 - 用户体验至上!

□ 青铜时代

2003年，非典肆虐京城，中关村所有实体店都是门可罗雀，京东也不例外。当时，老刘为了员工的安全，决定关闭所有门店应对非典。可门店关闭后，京东的资金最多能支撑半年，公司如何经营是一个大问题。偶然的机会，老刘发现通过网上论坛可以直接向客户销售商品，而且不必与客户接触，能避免非典的影响。就是这样，京东开始了网络销售之路。

经过一段时间的尝试，线上销售额持续增长，网络销售为公司打开了新天地。老刘经过慎重考虑决定关闭线下实体店，专营网上销售。而网络销售不能单单依靠公用的论坛，还是要建立京东自己的线上销售平台。这个平台将成为京东未来开展业务的最主要渠道，它也被誉为京东的生命线。



2003年老刘打算自建网站的时候就想到了曹鹏，当时除了曹鹏还邀请了另外一个程序员。据曹鹏回忆，当时就他们两个人开发这个网站，用ASP技术。

2004年初，网站正式上线，名为“京东多媒体网”。网站的功能比较简单实现了商品陈列，用户可以顺利下单。而下单后，则需要人工打印单子到库房提货，操作比较原始。无论如何，这是京东第一个自主研发的网络销售平台，是京东电商事业的重要里程碑。利用这个平台，用户实现了自主线上购买，京东与用户的距离更近了一步。网站上线后，销售额不断扩大，增长速度惊人。

从2004年到2006年，负责网站开发和管理的人员不足5人。所有与系统相关的工作都由这几个人合力完成。从论坛到后来自建网站，老刘对用户反馈异常关注，他也一直保持着与用户密切沟通、回复用户留言的习惯。因此，老刘特别重视网站的开发工作。据老员工回忆，当时他在公司有两处工位，其中之一就在研发部。平时工作中，他也经常指导网站的设计、文案的撰写等。

2007年8月京东拿到第一笔风险投资，企业发展进入快车道。订单量的增长也对系统提出了巨大挑战。京东急需找到一位优秀的技术人才来领导团队，为业务增长打好基础。

2008年5月，Donny加入京东，担任京东研发部负责人。进入京东后，Donny仔细研究了现有系统，认为现有系统无法满足京东未来的业务增长，必须进行架构升级，从ASP转向.NET。同时，借此机会对前台页面进行改版，提升用户体验。



□ 白银时代

2008年8月20日，一辆面包车停在了北京通州区某个高档小区的一栋三层别墅门前，从车上下来了3个年轻人，他们好奇地四周张望着。别墅里的人很快出来了，一边向新来的年轻人打招呼，一边从面包车上卸下几台电脑和显示器，还有几个路由器和插线板，有说有笑地搬进了屋。京东未来的网站架构，就诞生在这座三层别墅当中。

在3个年轻人中，有一个刚刚入职3天的小伙子Frank，他从搜狐离职来到京东，希望在电子商务大潮中施展自己的才华，不过没想到的是，加入京东后的第一项任务，就是收拾行囊投入为期将近3个月的封闭开发。

3个月后，经过大家的努力拼搏，封闭开发顺利结束了。新版京东商城网站在11月初顺利上线。在与用户互动的过程中，虽然发现了一些Bug和性能问题，但经过修复，用户总体满意度和转化率还是有比较大提升的。在这次改版中，网站页面比较大的改动有4处：第一是商品分类不再采用全部罗列的方式，而是改为二级分类弹层模式；第二是增加商品排行榜；第三是弱化了登录，节省空间展示更多商品；第四是购物车展示在首页，方便用户随时结算。

新版网站为京东在2008—2010年的快速发展打下了坚实基础，订单量从日均几千单增长到十几万单。同时，经过3个月的磨合，一个崭新的研发团队也随即诞生。这支队伍的成员也都逐渐成长为京东技术的骨干力量。



□ 黄金时代

2010年在老刘的提议下，负责订单展示的产品经理开始着手设计“订单跟踪”功能。这个功能很简单，就是把从用户下单开始的整个服务环节的关键节点都记录并显示出来，无论用户何时打开网页，都能看到自己的商品现在在哪，是什么状态，哪个配送员负责，他的电话是多少。

从页面内容来看，这是一个非常简单的信息展示，但其背后的实现机制并不简单，这需要从用户下单那一刻起，京东信息系统就把商品的所有流转都记录下来，这需要订单系统、仓储系统、分拨系统、配送系统、站点管理系统、配送员信息系统等各个系统的打通和联动，才能最终将每一条信息完整、准确地显示给用户。

京东做到了。订单跟踪系统刚刚上线，就获得了用户的一致认可，客服电话呼入率大幅降低，用户满意度大幅提高。这种“一切尽在掌握”的感觉对用户来说是极好的。在电商需要被更多人接受和信任的年代，这种服务的确直击痛点，并很快成为各大电商效仿的对象。虽然现在“订单跟踪”已经成为电商的标配，但京东是最先推出该功能的。这种对用户体验的关注和创新精神，一直推动着京东的快速前进。





京东 2011 年系统升级架构简图

到了 2012 年，随着仓储配送系统的加强、流程制度的完善及大数据的积累，京东进而推出订单时效 Promise 系统——根据商品库存状态、仓库所在地、收货地、下单时间、配送能力等因素，提前预测订单送达时间，让用户更早地知道何时能收到货。比如，在周五上午 11 点前下单，北京的用户会看到“预计下午 6 点前送达”的提示，这样用户就可以放心在单位收货，而不用纠结到底是送到单位还是送到家里。订单 Promise 从服务角度来讲，是京东给用户的购物承诺，当有送货延迟，会给用户一定补偿。同时大数据团队也会反复测算数据，找到问题出在哪里，如何更精准的预测。京东靠着对自己严苛的要求，不断提升着用户体验。“京东送货快”，就是这样一点一滴形成用户认知和良好口碑的。

不光是在物流服务中，京东利用大数据的分析方法。在用户浏览网站时，我们也要通过数据分析用户的习惯，从而指导网站设计。而如何收集用户浏览习惯的真实数据呢？



建立用户体验室

2006年年底，一本书悄然在互联网圈，尤其是产品设计圈流行起来，它就是 *Don't make me think*。此后，用户体验更多地被重视和传播开来，什么是好的用户体验，怎样做出好的用户体验，每个互联网公司都在思考，京东也不例外。

做好用户体验，就是要多和用户接触。老刘在这方面已经给京东的产品经理树立好榜样，经常亲自参与网友见面会、接客服电话、在线交流等。而要想和用户在一起，首先必须能快速找到用户。对此，京东的产品经理们马上想到一个地方——京东自提点。在京东发展的初期，用户还抱有试试看的心态在京东购物，一手拿货一手交钱的习惯让不少人选择了来京东自提点先提货后付款的购物方式。这里有新用户也有老用户，不同性别、年龄的用户都可以找到，而且地点固定、客流量较大，所以在早期，京东自提点就是京东产品经理们做用户调研和访谈的最佳场所。

用户体验室的建立为京东在 2012 年的改版奠定了基础。2011 年下半年京东启动前台网页改版。本次改版的核心宗旨是展现京东综合商城的形象，同时提升美观性，吸引更多女性用户的关注。

新版网站上线后，通过数据对比发现，首页转化率提升了 4.74%。用户体验进一步改善。

别人的提炼 - 京东网站的技术架构

□ 前端采用.NET技术架构

- 说起京东商城，不得不提到它的掌门人，2011CCTV十大经济人物“刘强东”。
- 刘强东毕业于中国人民大学，理科超强的他业余时间自学编程。在创立京东商城的早期，刘强东选用了最为熟悉的.NET，直到现在，京东商城的前端部分仍然沿用.NET技术架构。

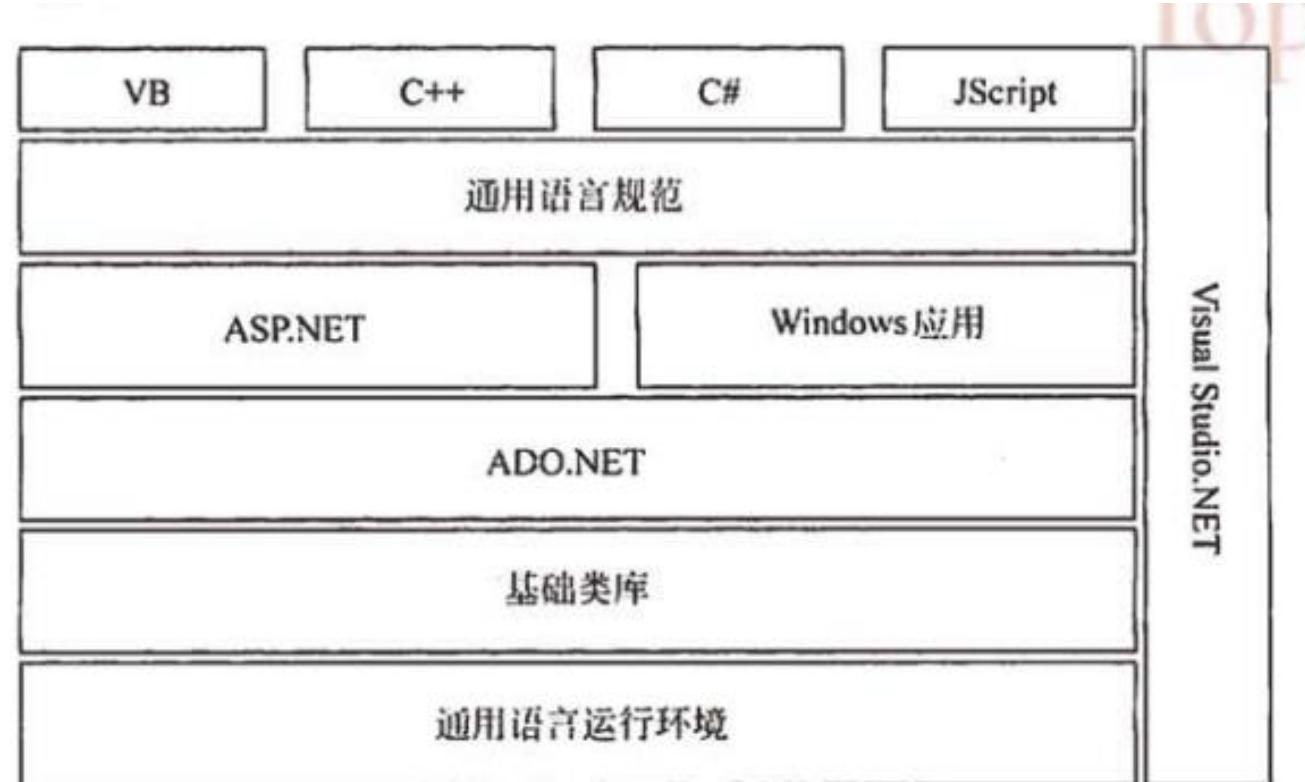
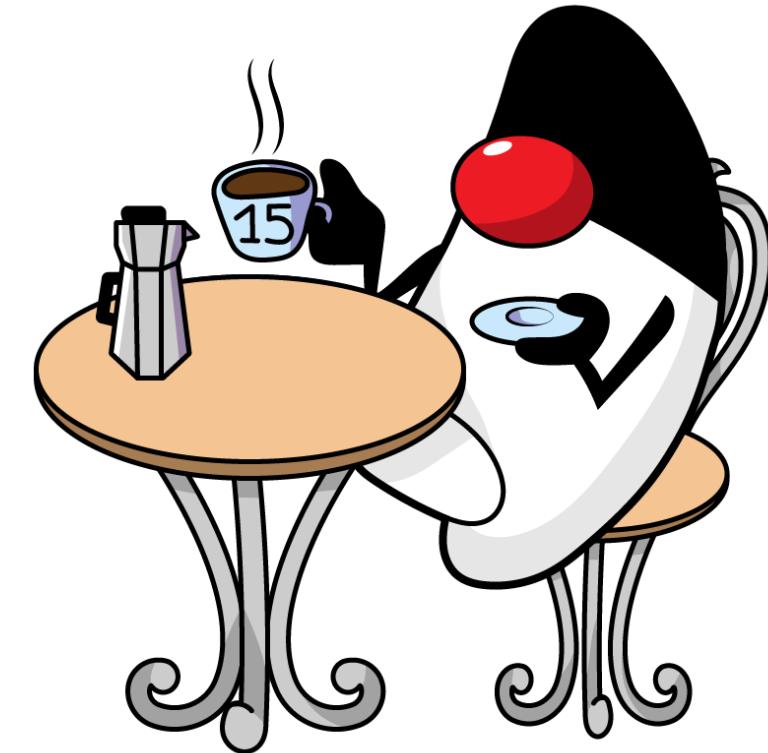


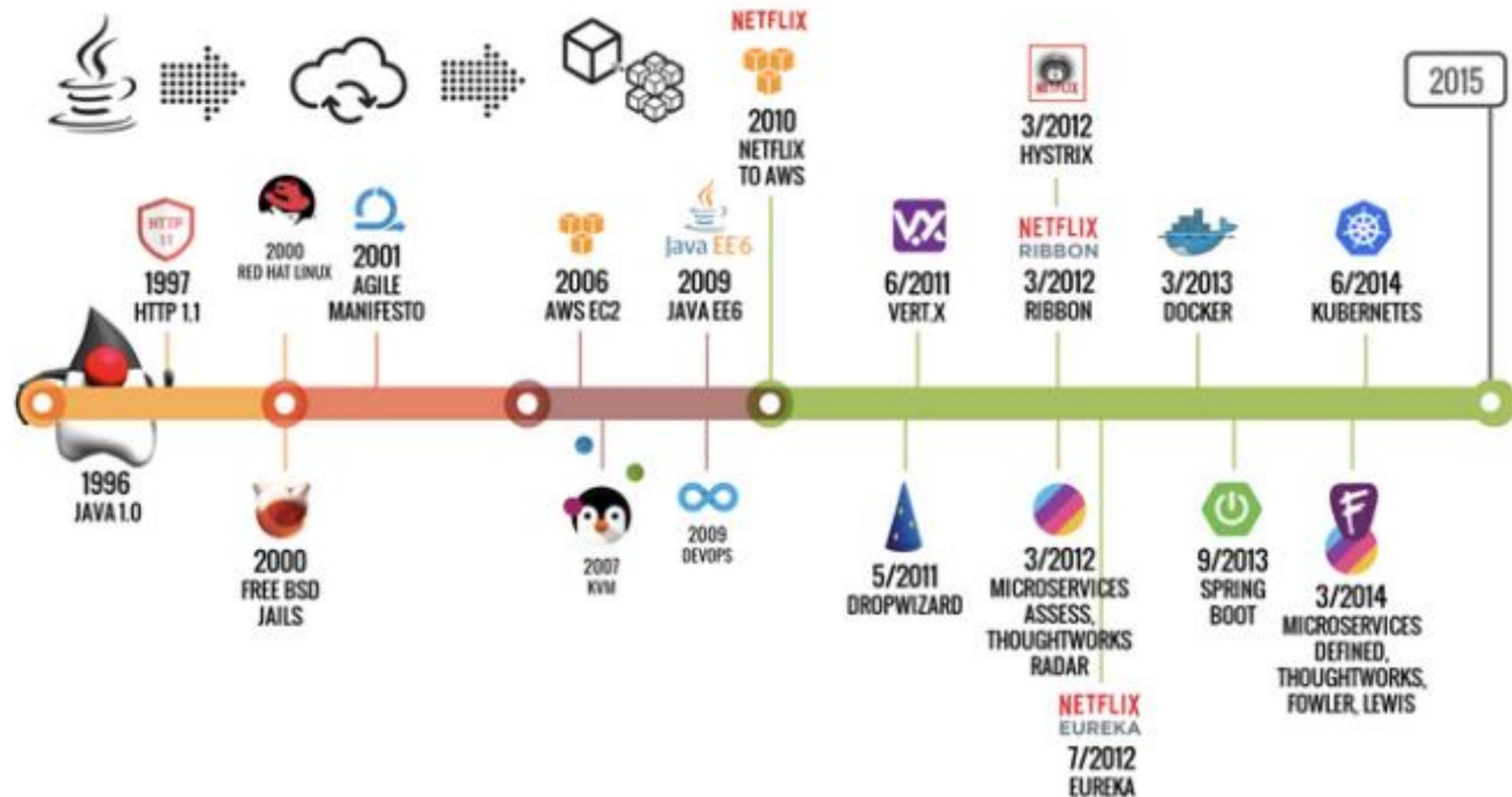
图 3-12 .NET 开发框架

□ 后端采用Java技术

- 京东商城连续六年200%以上的增长率，并不能掩盖其背后隐含的技术难题。
 - 比如，在2011年11月1日京东商城的图书促销活动中，由于促销活动仅限时1个小时，消费者疯狂抢购导致服务器不堪重负，最终瘫痪。为平息消费者的抱怨，刘强东在微博上表示：“重搞活动，增加3倍服务器，活动时间不能低于3小时。”尽管刘强东的危机公关做得很到位，即使京东商城增加了三倍服务器，技术问题依旧没有解决。
- 随后京东商城进行了技术构架调整，基于历史原因前端技术依然使用.NET技术开发并且进行了大规模的性能优化，**将系统后台使用JAVA技术进行重构，暂时性的缓解了系统访问压力。**



Timeline of technology contributions to Java



□ 团购采用PHP技术

■ 据团购导航网站团 800最新统计，2011年国内团购市场交易总额达110亿元，比2010年实现了550%的增长；超过3亿人次“抱团”消费，而按全年平均折扣3.6折计算，团购为全国消费者共节省了超过190亿元的消费支出，比2010年多省了156亿元。看看上述这组数字，就不难想象京东商城对于团购的蛋糕是多么的垂涎欲滴。

■ **与众多团购网站一样，京东商城的团购频道采用了高效率的PHP技术。**

➤ 与其它语言相比，PHP是将程序嵌入到HTML文档中去执行，执行效率比完全生成HTML标记的CGI要高许多；与同样是嵌入HTML文档的脚本语言 JavaScript相比，PHP在服务器端执行，充分利用了服务器的性能；PHP执行引擎还会将用户经常访问的PHP程序驻留在内存中，其他用户在一次访问这个程序时就不需要重新编译程序了，只要直接执行内存中的代码就可以了，这也是PHP高效率的体现之一。PHP具有非常强大的功能，所有的CGI或者JavaScript的功能PHP都能实现，而且支持几乎所有流行的数据库以及操作系统。



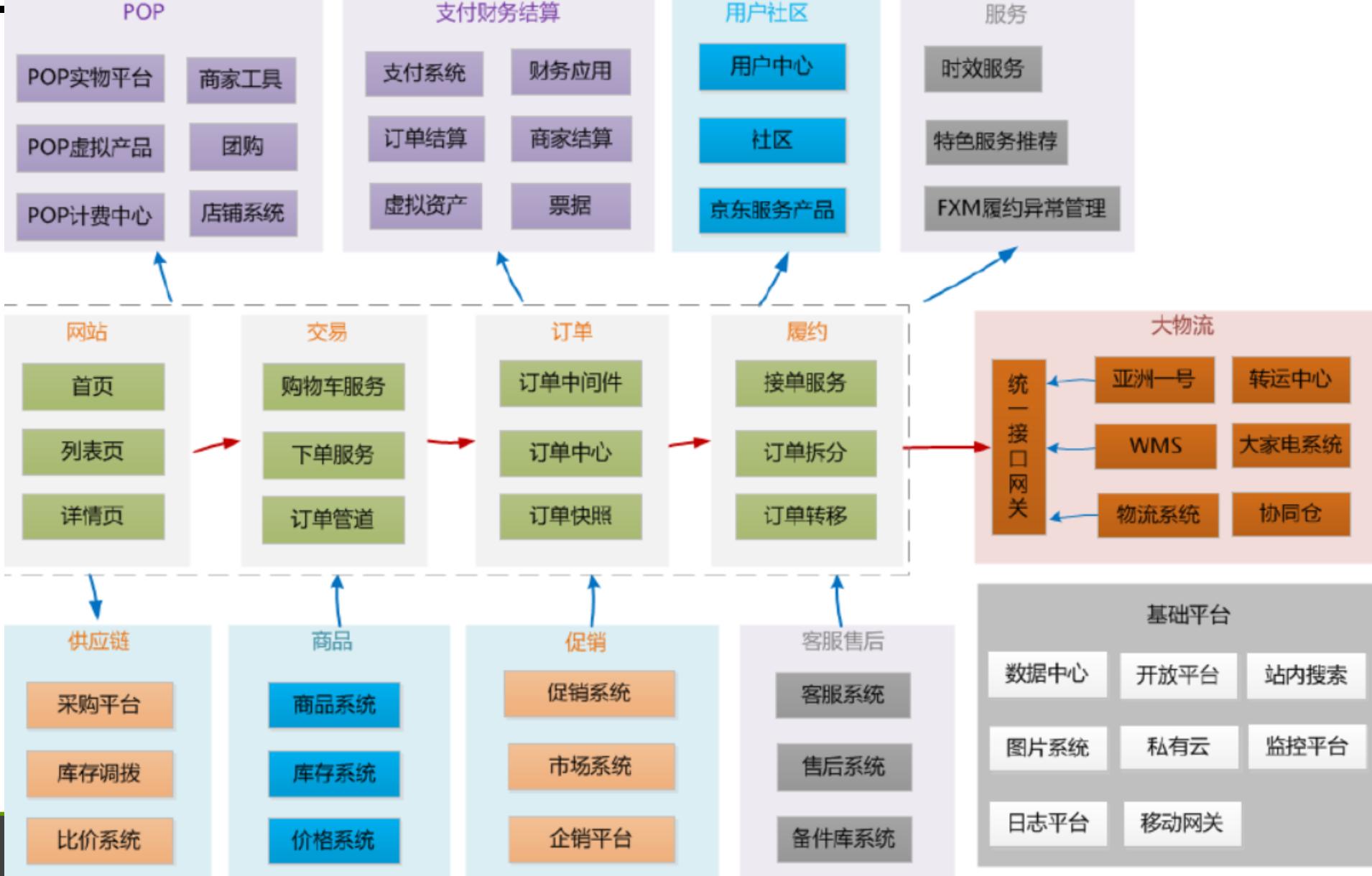
□ UI采用jQuery插件

- 习惯网购的朋友都深有体会。大部分皇冠级淘宝卖家都是图片控。京东商城的放大图效果也是吸引消费者的法宝之一。京东商城产品展示页支持多图切换并放大代码，**放大功能的核心代码为jqZoom**，当把鼠标移到下边的小图上的时候，上边显示大图片，当把鼠标放到大图片上，右侧则出现鼠标对应地方的大图片，方便查看产品细节，比较适合网店产品展示。
- JQZoom 是一个JavaScript图像放大镜，内置在流行的jQuery JavaScript框架顶部。jqZoom 功能非常强大，操作较为简单。支持标准模式、反转模式、无镜头、无标题的放大，并可以自定义jqZoom的窗口位置和渐隐效果。

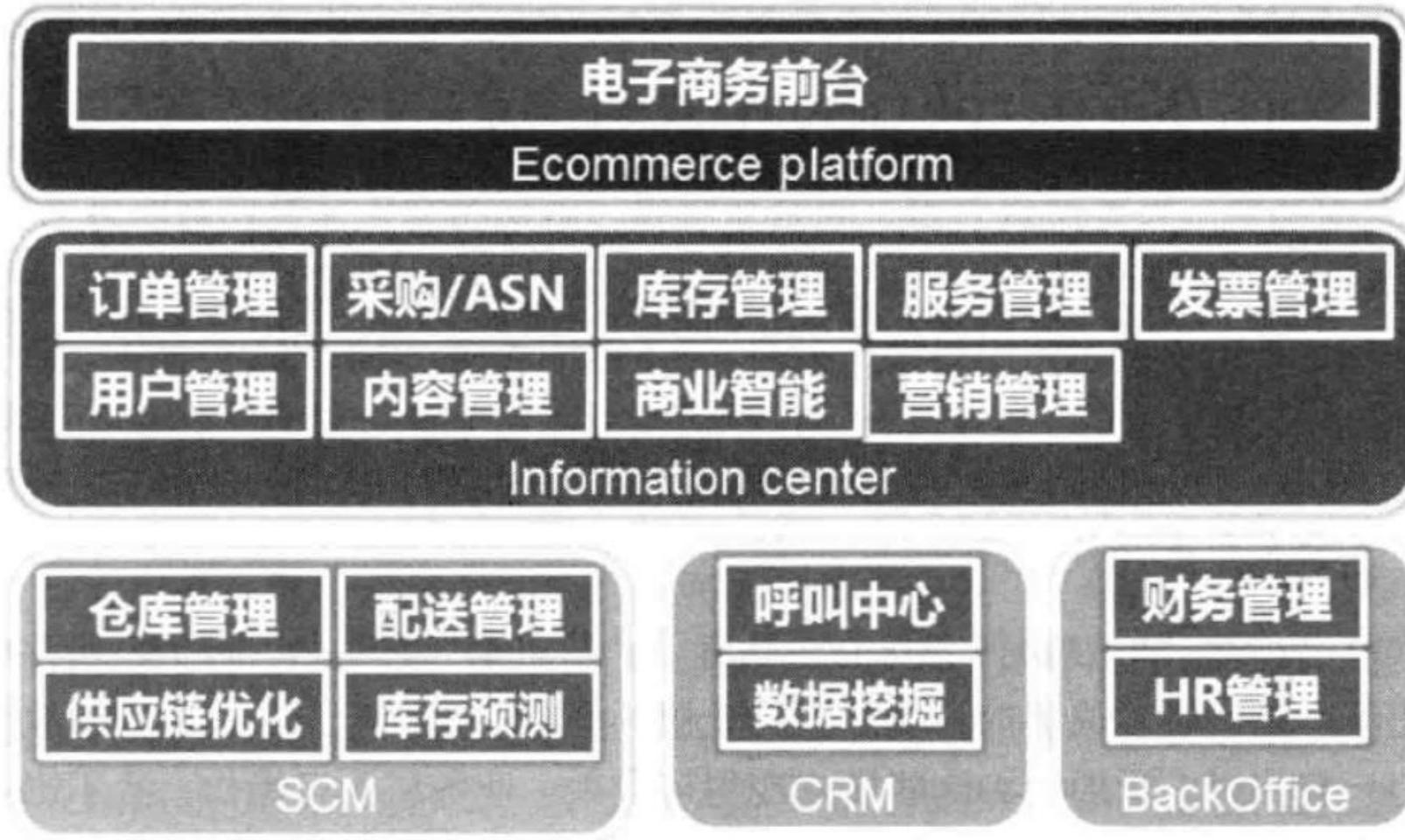


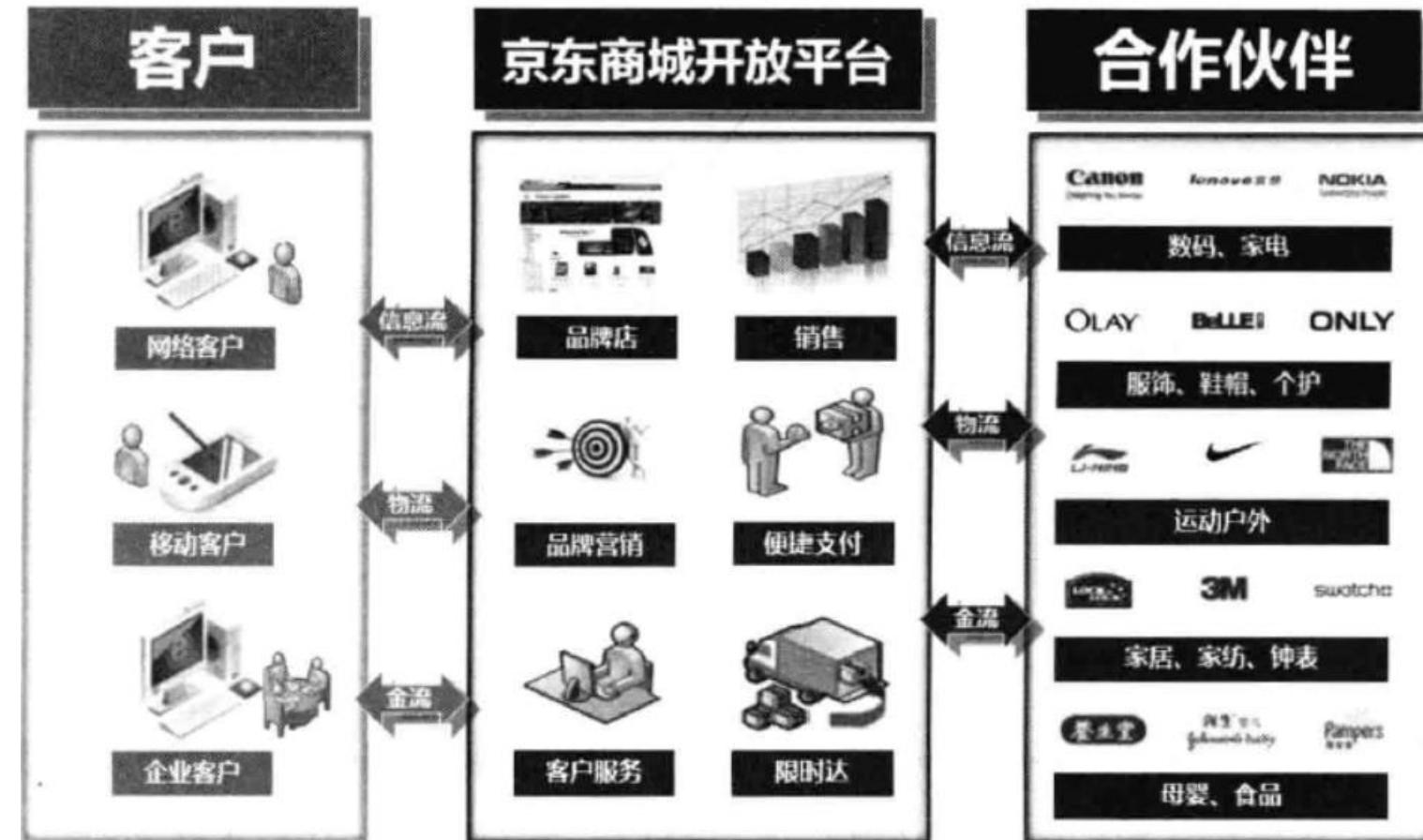
业务架构

POP: Platform Open Plan



□ POP: Platform Open Plan





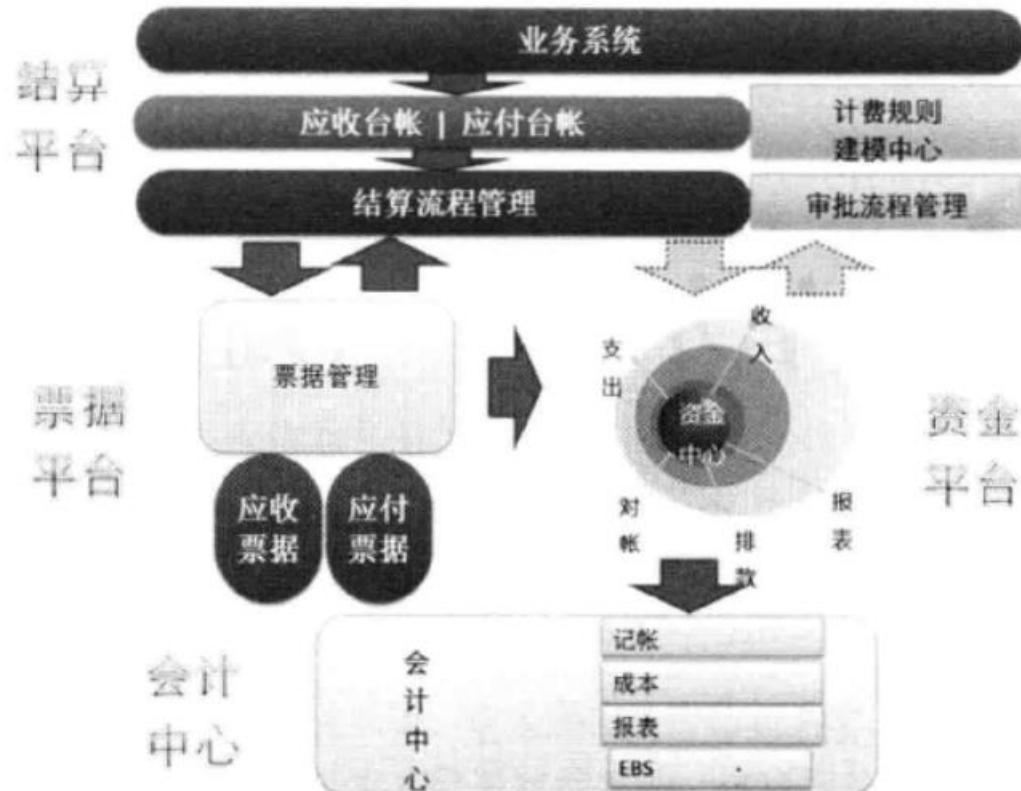
□ POP平台下共分为四大模式——FBP/LBP/SOPL/SOP

- 1、FBP (Fulfillment By POP) :卖家在京东销售商品，京东提供仓储来管理所销售商品，京东完成购物订单配送和收款，京东开具发票给消费者。
- 2、LBP (Logistics By POP): 卖家在京东销售商品，卖家每日将消费者订单打包送京东分拣中心，京东完成购物订单配送和收款，京东开具发票给消费者。
- 3、Sopl(Sale On POP&Logistics By POP): 卖家在京东销售商品，卖家每日将消费者订单打包送京东分拣中心，京东完成购物订单配送和收款，卖家开发票给消费者。
- 4、SOP(Sale On pop): 卖家在京东销售商品，卖家每日将消费者订单打包并自行或采用快递完成购物订单配送，卖家开发票给消费者。



POP 系统功能框架

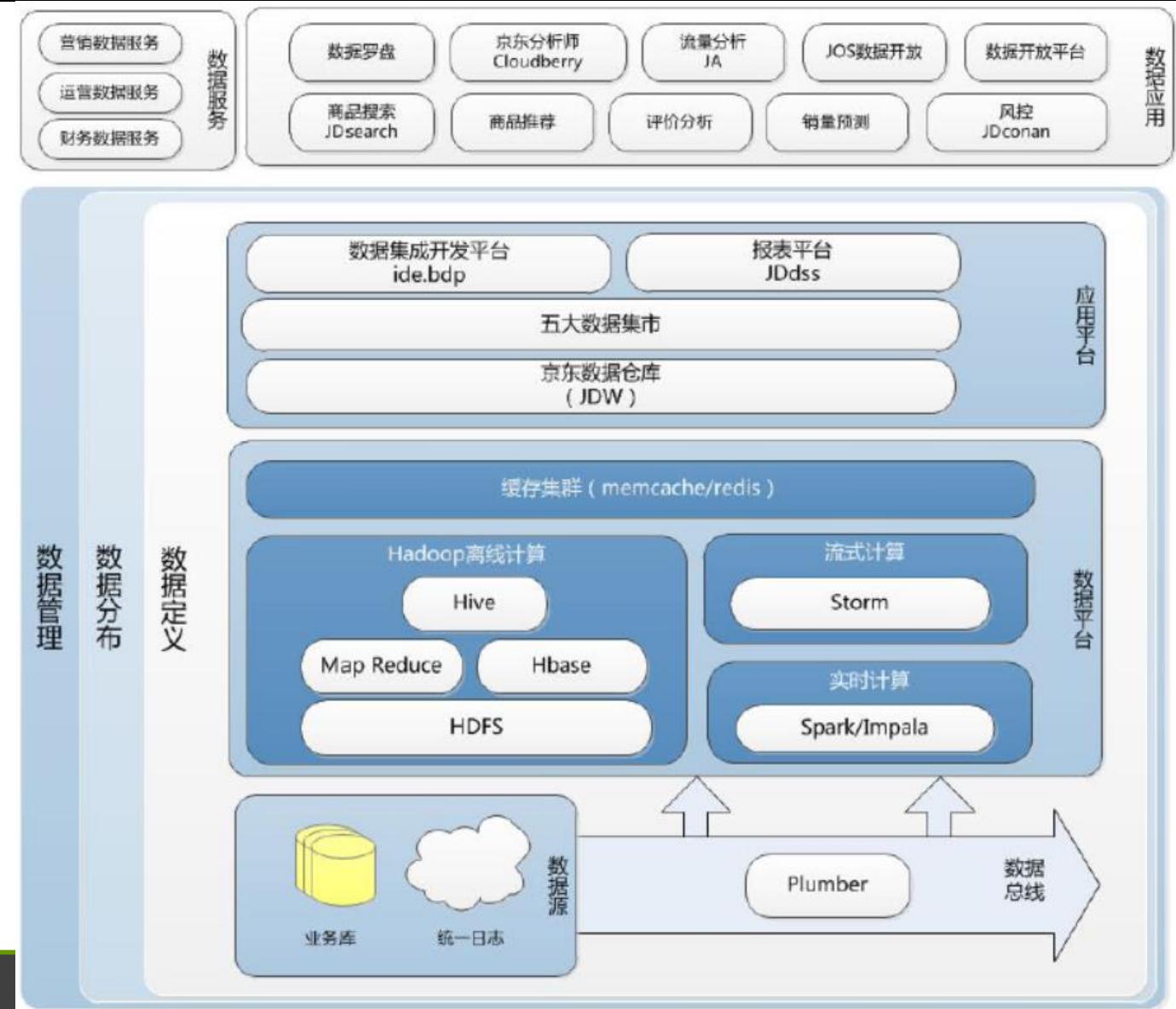
□ 京东财务



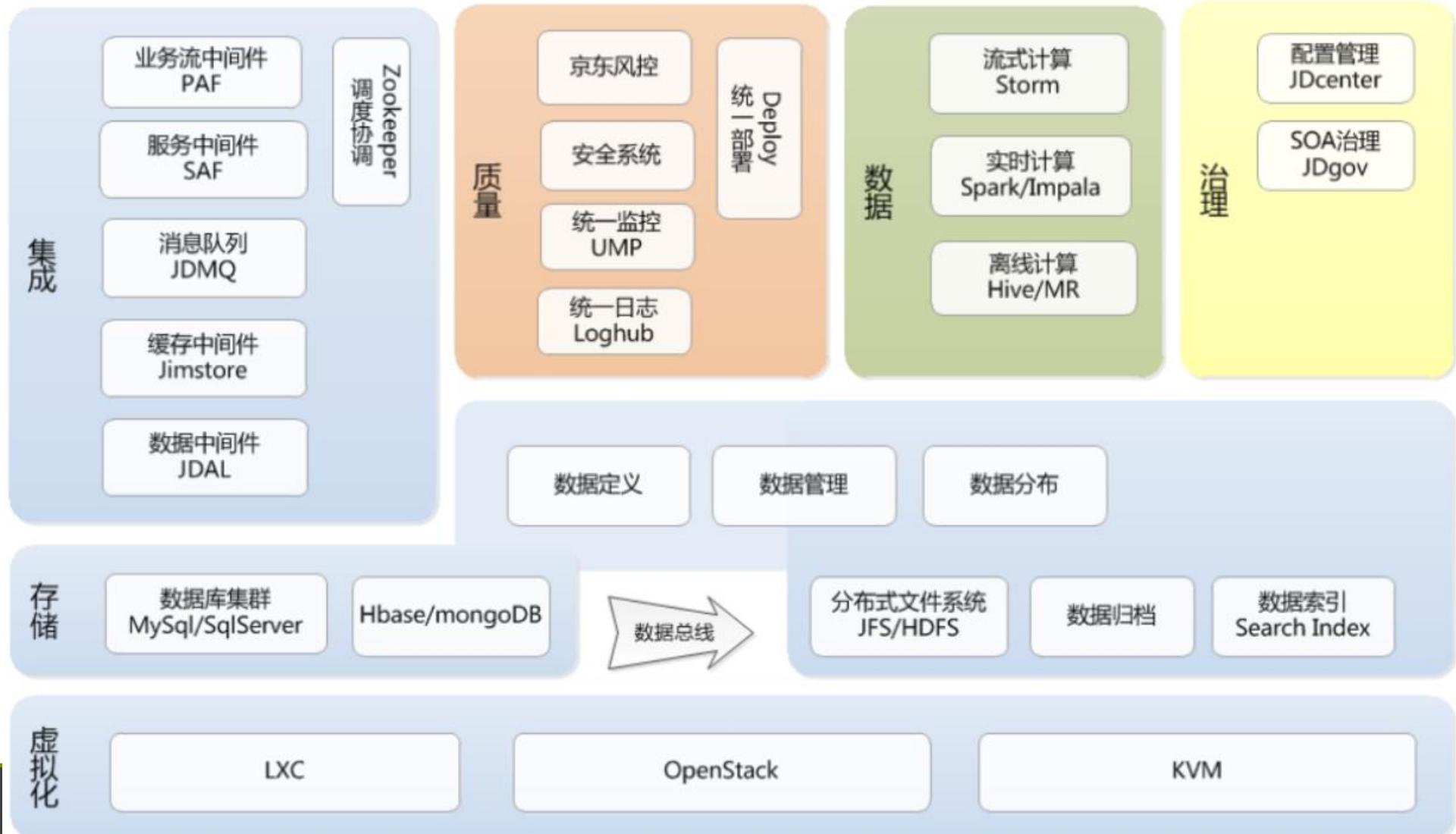
京东财务体系



□ 数据架构



□ 基础架构

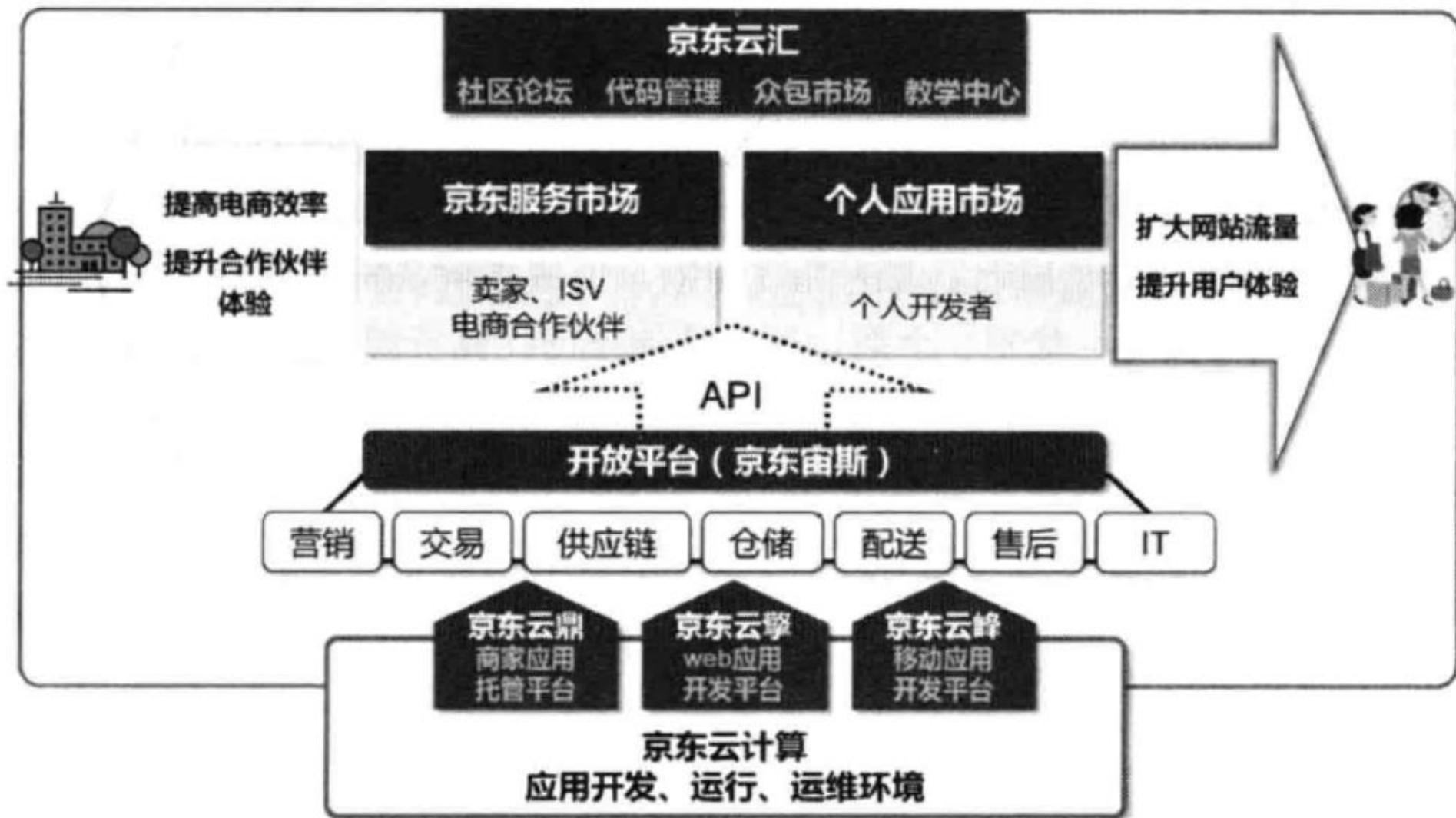


新技术~~

□ 京东云



京东云整体架构图



新技术~~

试衣



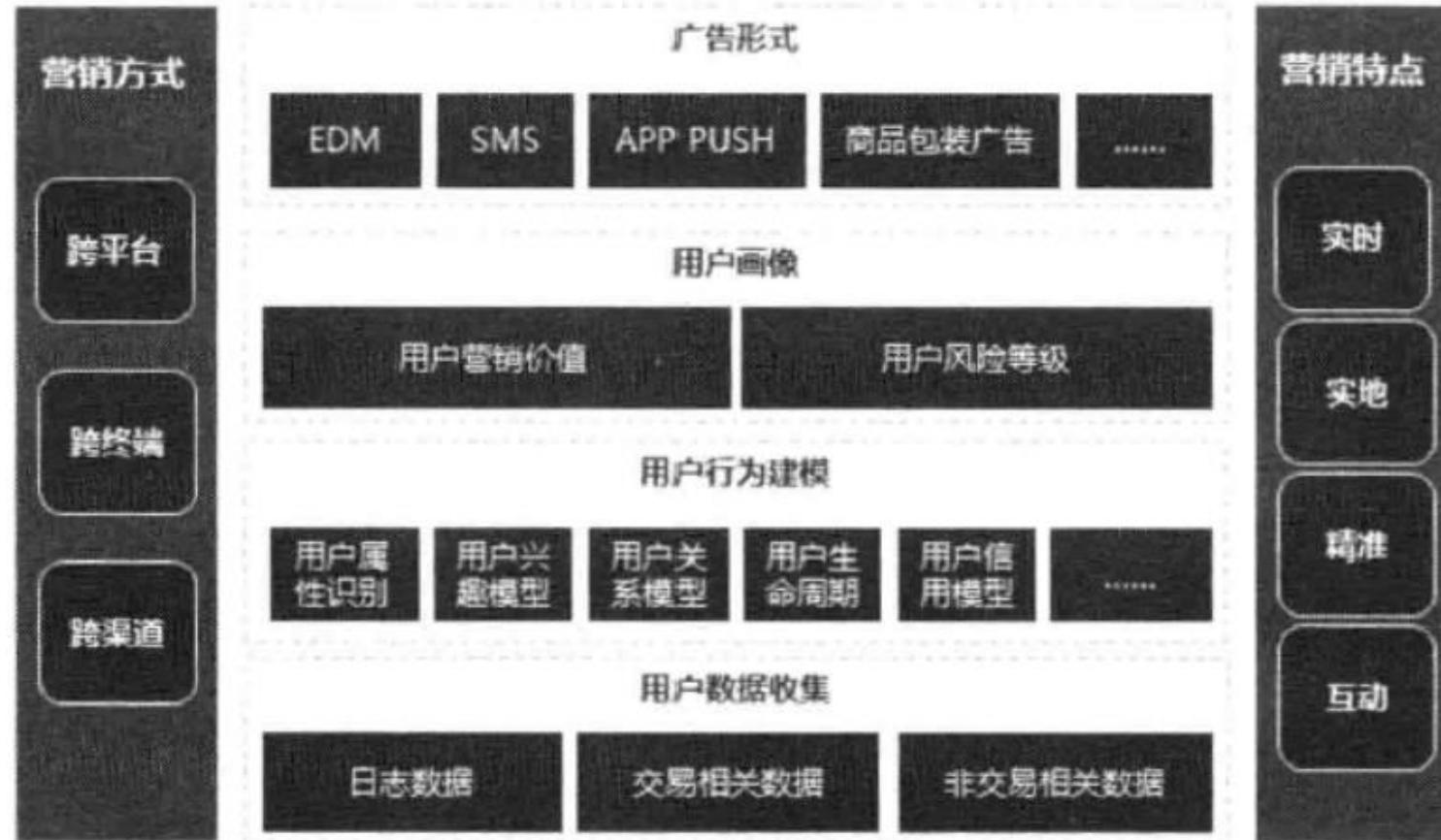
京东试衣间



模特试穿图

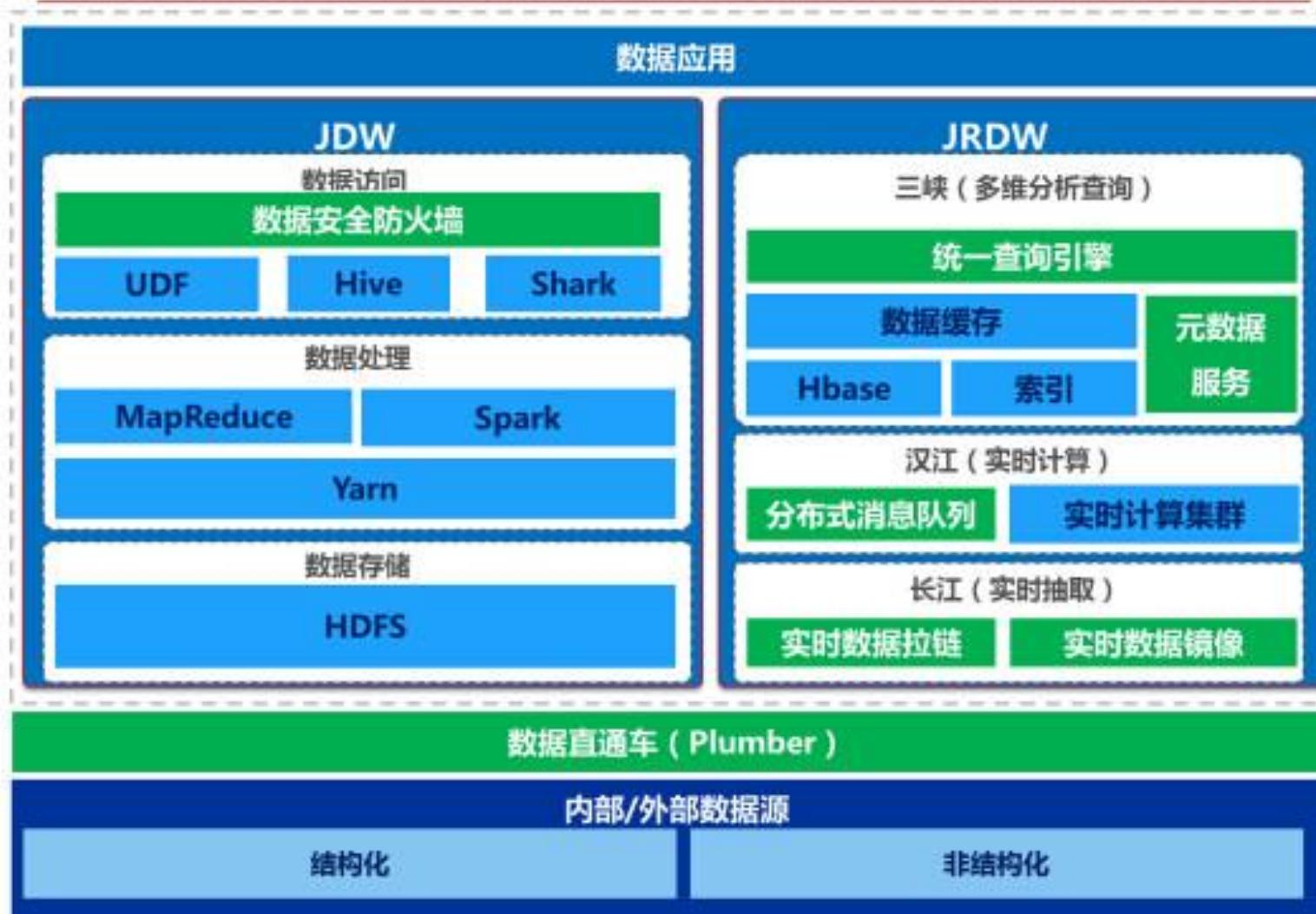
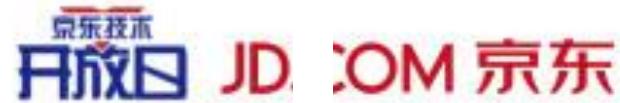
新技术~~

□ 大数据分析+人工智能



精准营销架构

京东大数据平台



JDMP-支持的算法



推荐

ALS-MF , FP-Growth , Item/User-CF , RBM

分类

LR , NB , SVM , gbdt , soft-max

回归

linear , ridge , lasso

聚类

k-means

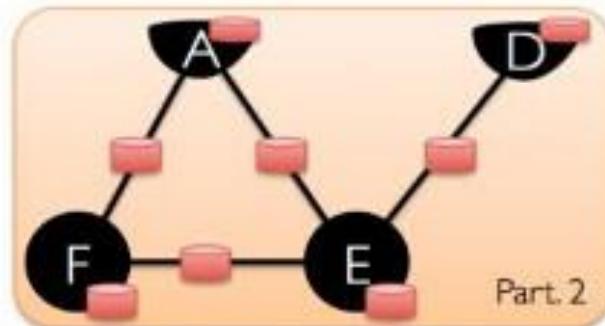
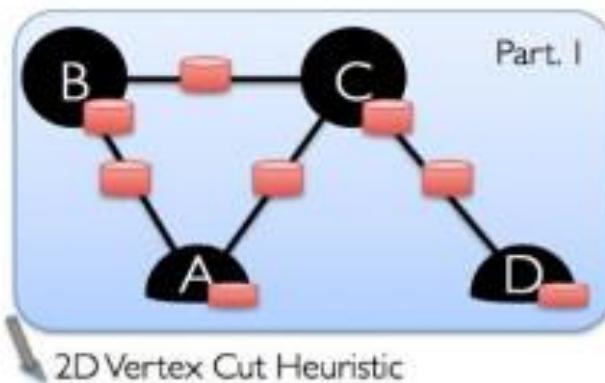
主题模型

LDA , PLSA

JDMP-Graphx图计算



Property Graph



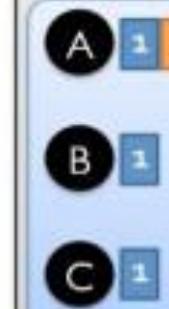
Vertex Table
(RDD)

Routing
Table
(RDD)

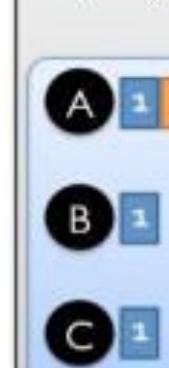
Edge Table
(RDD)



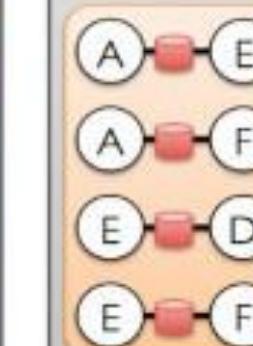
Routing
Table
(RDD)

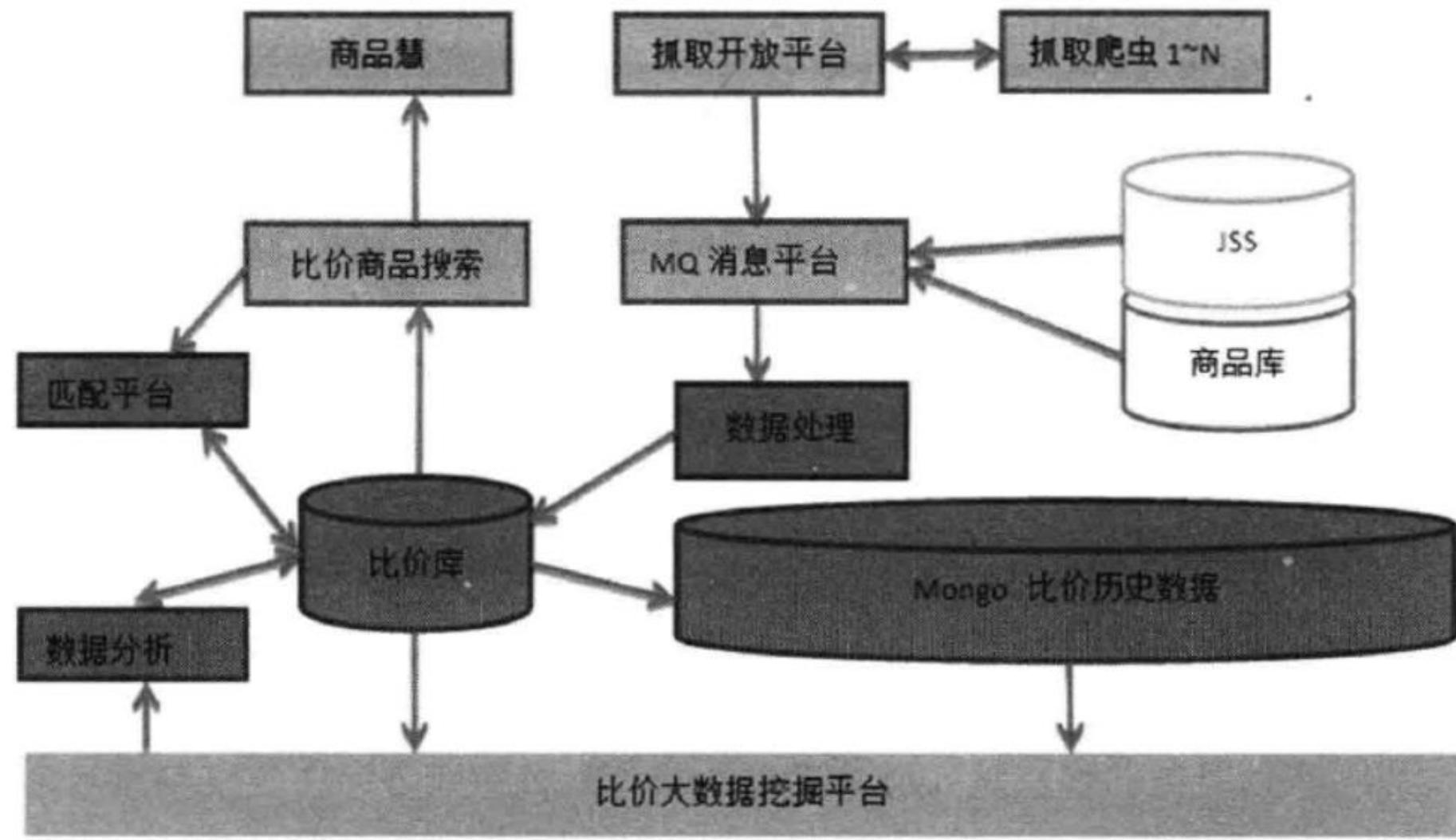


Edge Table
(RDD)



Edge Table
(RDD)





大数据环境下比价系统架构图

销量预测基本逻辑

JD.COM 京东



京东数据集市

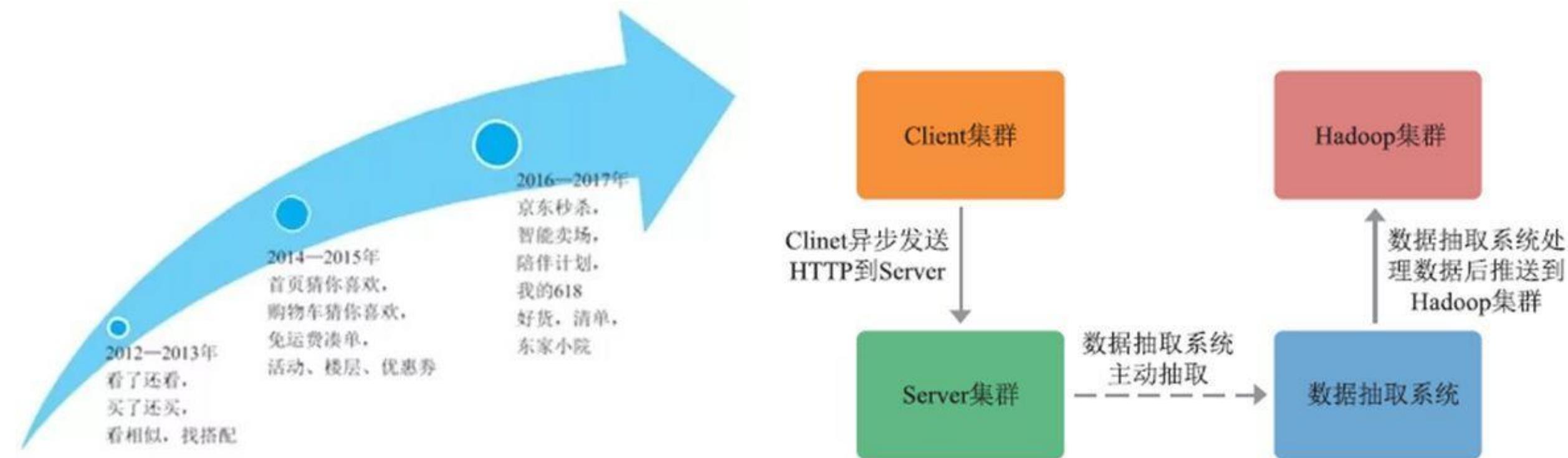


销量预测自动补货库存健康系统简介

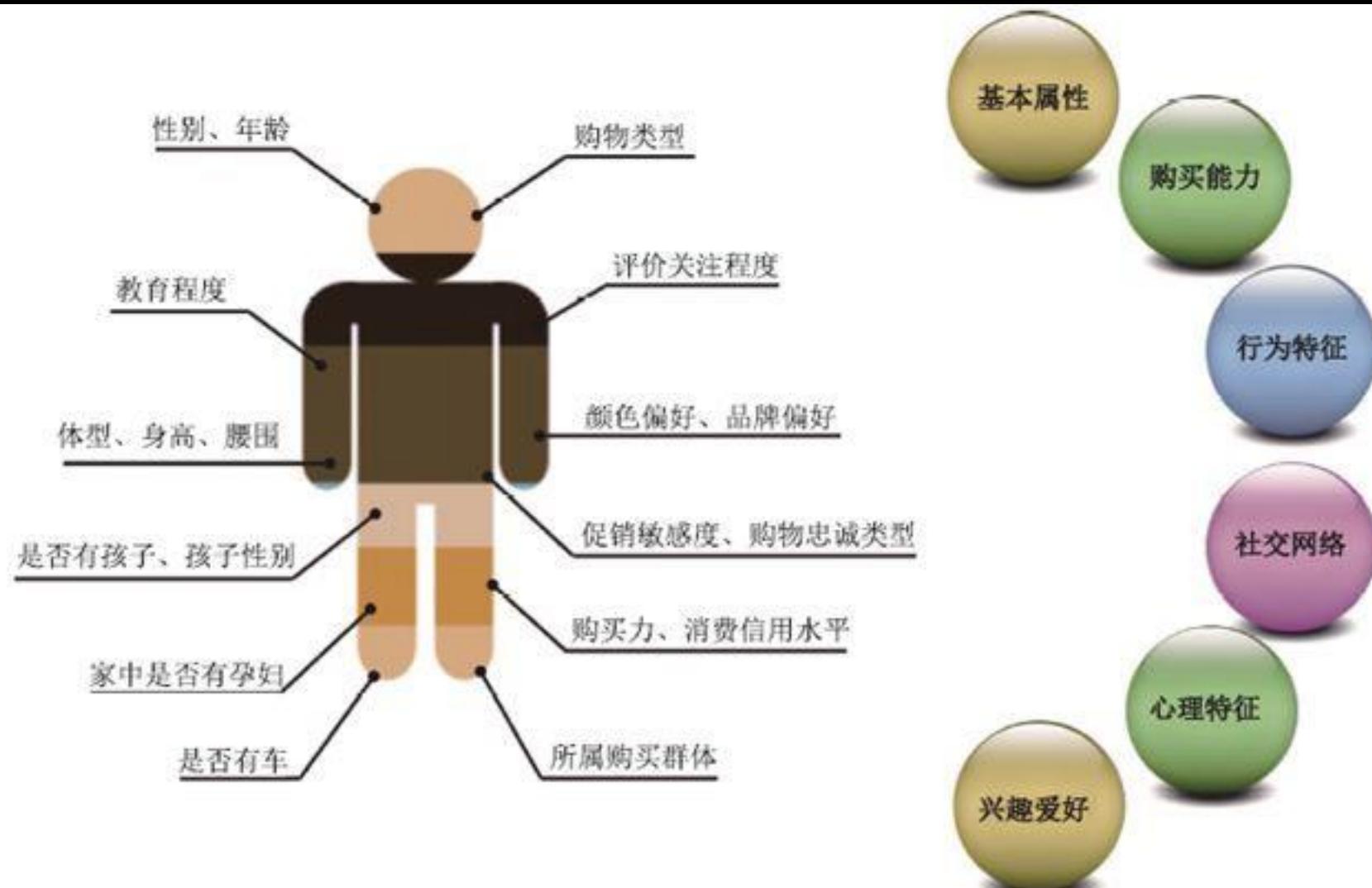
JD.COM 京东



□ 京东个性化推荐系统







欢迎



营销策略

微信绑定

精选标签

中青年女士
34岁
股龄3年
积极型(C4)
低潜力
东莞
是否有效户:否

股票操作频繁度:C.静若处子型
是否一人多户:一人多户
APP活跃度:最近一月登陆2次
开户营业部:3210
首次领用投顾VIP:否
个股期权开通:否
非货基破冰:否
创业板开通:否
港股通开立:否
同业是否开两融:否
最小同业市值:0
20日均资产50万:否
两融适格:否
客户类型:综合型
客户在其他券商资产:1万~5万
中登开基金户年限:1年

重要事件提醒

问卷调查
KYC问卷

账户开立+产品破冰 **客户投资状况**

首次登陆APP
绑定微信
购买新客专享产品
购买活钱宝
首次银证转账
二选一
资金账户
理财账户
沪A股东户
深A股东户
转签创业板
两融账户

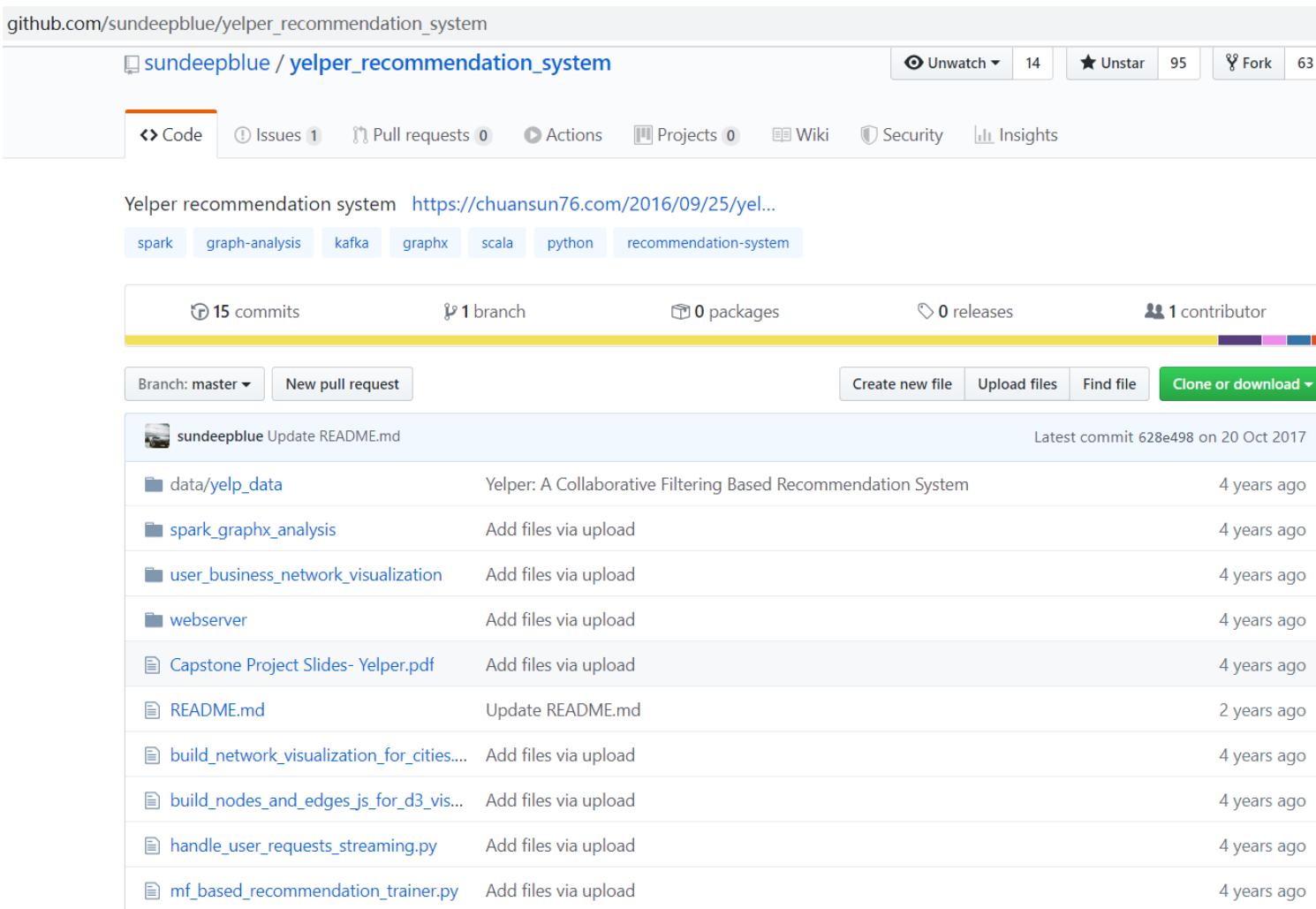
走势图 **平台成长记录**

● 总资产 ● 股票市值 ● 理财保有量



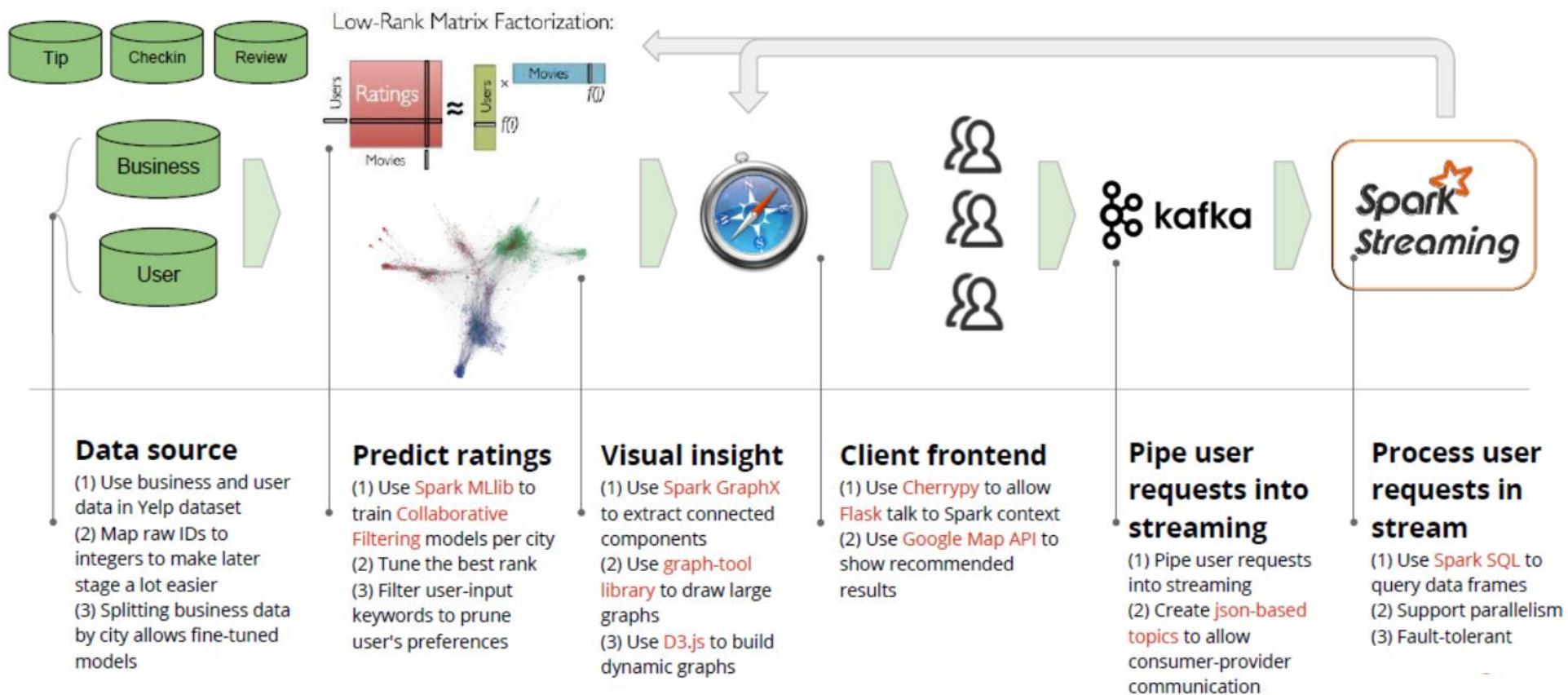

Yelper: A Collaborative Filtering Based Recommendation System

github.com/sundeepblue/yelper_recommendation_system

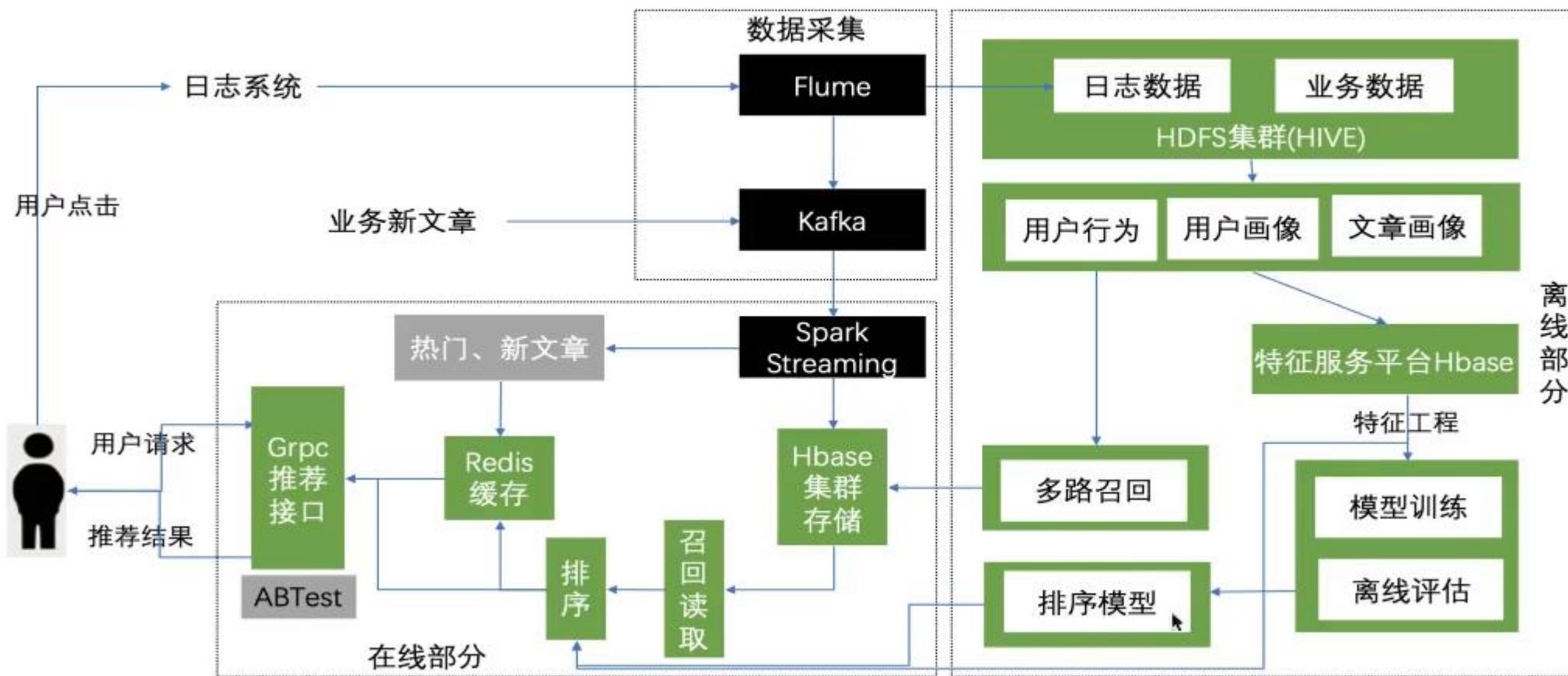
A screenshot of a GitHub repository page for 'yelper_recommendation_system' by user 'sundeepblue'. The page shows basic repository statistics: 15 commits, 1 branch, 0 packages, 0 releases, and 1 contributor. It includes a commit history table with details like file changes, commit messages, and timestamps. The repository is described as a 'Collaborative Filtering Based Recommendation System'.

| File / Action | Description | Time |
|--|---|--------------------------------------|
| README.md | Update README.md | Latest commit 628e498 on 20 Oct 2017 |
| data/yelp_data | Yelper: A Collaborative Filtering Based Recommendation System | 4 years ago |
| spark_graphx_analysis | Add files via upload | 4 years ago |
| user_business_network_visualization | Add files via upload | 4 years ago |
| webserver | Add files via upload | 4 years ago |
| Capstone Project Slides- Yelper.pdf | Add files via upload | 4 years ago |
| README.md | Update README.md | 2 years ago |
| build_network_visualization_for_cities.... | Add files via upload | 4 years ago |
| build_nodes_and_edges_js_for_d3_vis... | Add files via upload | 4 years ago |
| handle_user_requests_streaming.py | Add files via upload | 4 years ago |
| mf_based_recommendation_trainer.py | Add files via upload | 4 years ago |

Yelper is built on top of Spark platform



3.3.1.1 黑马召回与排序业务流程



Chapter 8: Case Study

□ History of JD

□ Overview of used techniques in JD

□ Architecture

□ Software architect [软件架构师]

大型互联网应用系统的特点

与传统企业应用系统相比，大型互联网应用系统有以下特点。

高并发，大流量：需要面对高并发用户，大流量访问。Google 日均 PV 数 35 亿，日均 IP 访问数 3 亿；腾讯 QQ 的最大在线用户数 1.4 亿（2011 年数据）；淘宝 2012 年“双十一”活动一天交易额超过 191 亿，活动开始第一分钟独立访问用户达 1000 万。

高可用：系统 7×24 小时不间断服务。大型互联网站的宕机事件通常会成为新闻焦点，例如 2010 年百度域名被黑客劫持导致不能访问，成为重大新闻热点。

海量数据：需要存储、管理海量数据，需要使用大量服务器。Facebook 每周上传的照片数目接近 10 亿，百度收录的网页数目有数百亿，Google 有近百万台服务器为全球用户提供服务。

用户分布广泛，网络情况复杂：许多大型互联网都是为全球用户提供服务的，用户分布范围广，各地网络情况千差万别。在国内，还有各个运营商网络互通难的问题。而中美光缆的数次故障，也让一些对国外用户依赖较大的网站不得不考虑在海外建立数据中心。



安全环境恶劣：由于互联网的开放性，使得互联网站更容易受到攻击，大型网站几乎每天都会被黑客攻击。2011 年国内多个重要网站泄露用户密码，让普通用户也直面一次互联网安全问题。

需求快速变更，发布频繁：和传统软件的版本发布频率不同，互联网产品为快速适应市场，满足用户需求，其产品发布频率是极高的。Office 的产品版本以年为单位发布，而一般大型网站的产品每周都有新版本发布上线，至于中小型网站的发布就更频繁了，有时候一天会发布几十次。

渐进式发展：与传统软件产品或企业应用系统一开始规划好全部的功能和非功能需求不同，几乎所有的大型互联网站都是从一个小网站开始，渐进地发展起来的。Facebook 是伯克扎克同学在哈佛大学的宿舍里开发的；Google 的第一台服务器部署在斯坦福大学的实验室里；阿里巴巴则是在马云家的客厅里诞生的。好的互联网产品都是慢慢运营出来的，不是一开始就开发好的，这也正好与网站架构的发展演化过程对应。

1.2.1 初始阶段的网站架构

大型网站都是从小型网站发展而来，网站架构也是一样，是从小型网站架构逐步演化而来。小型网站最开始时没有太多人访问，只需要一台服务器就绰绰有余，这时的网站架构如图 1.1 所示。

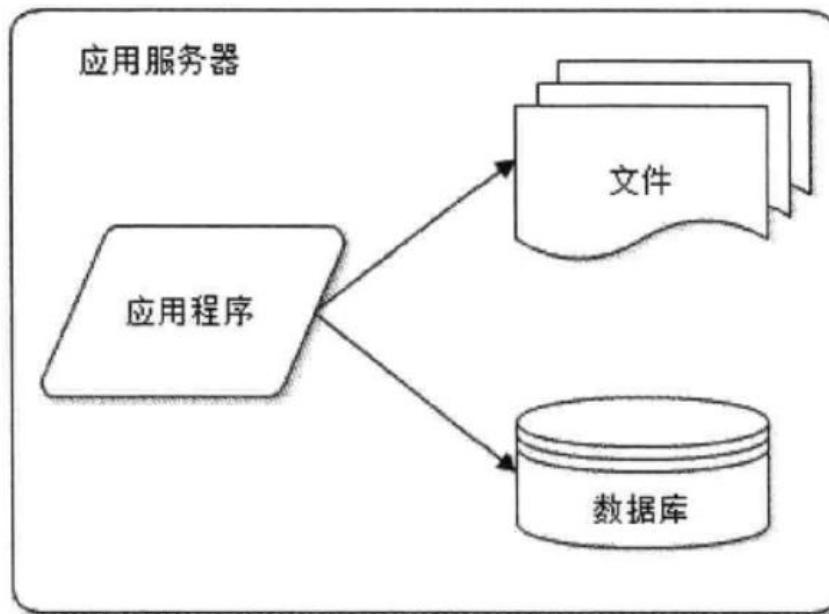


图 1.1 初始阶段的网站架构

应用程序、数据库、文件等所有的资源都在一台服务器上。通常服务器操作系统使用 Linux，应用程序使用 PHP 开发，然后部署在 Apache 上，数据库使用 MySQL，汇集各种免费开源软件及一台廉价服务器就可以开始网站的发展之路了。

1.2.2 应用服务和数据服务分离

随着网站业务的发展，一台服务器逐渐不能满足需求：越来越多的用户访问导致性能越来越差，越来越多的数据导致存储空间不足。这时就需要将应用和数据分离。应用和数据分离后整个网站使用三台服务器：应用服务器、文件服务器和数据库服务器，如图 1.2 所示。这三台服务器对硬件资源的要求各不相同，应用服务器需要处理大量的业务逻辑，因此需要更快更强大的 CPU；数据库服务器需要快速磁盘检索和数据缓存，因此需要更快的硬盘和更大的内存；文件服务器需要存储大量用户上传的文件，因此需要更大的硬盘。

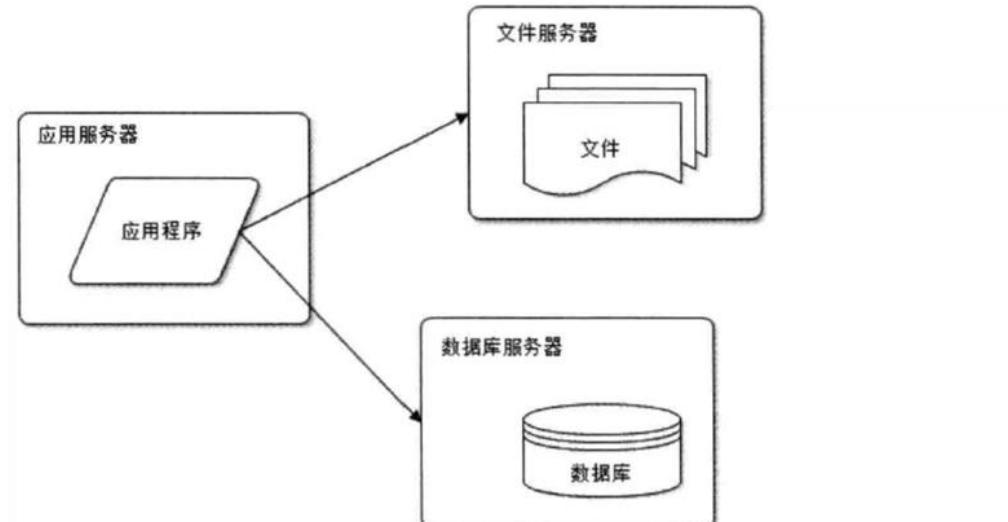


图 1.2 应用服务和数据服务分离

1.2.3 使用缓存改善网站性能

网站访问特点和现实世界的财富分配一样遵循二八定律：80%的业务访问集中在 20% 的数据上。淘宝买家浏览的商品集中在少部分成交数多、评价良好的商品上；百度搜索关键词集中在少部分热门词汇上；只有经常登录的用户才会发微博、看微博，而这部分用户也只占总用户数目的一小部分。

既然大部分的业务访问集中在一小部分数据上，那么如果把这一小部分数据缓存在内存中，是不是就可以减少数据库的访问压力，提高整个网站的数据访问速度，改善数据库的写入性能了呢？



网站使用的缓存可以分为两种：缓存在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。本地缓存的访问速度更快一些，但是受应用服务器内存限制，其缓存数据量有限，而且会出现和应用程序争用内存的情况。远程分布式缓存可以使用集群的方式，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，如图 1.3 所示。

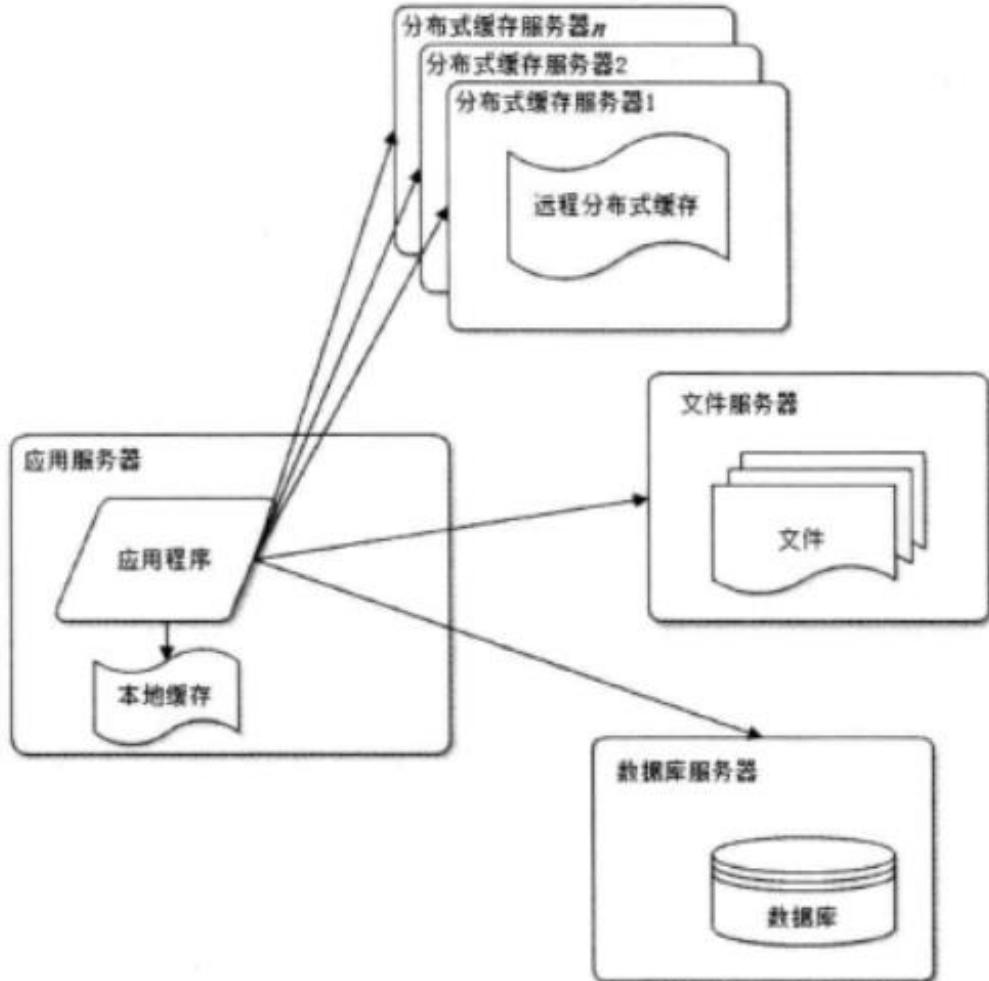
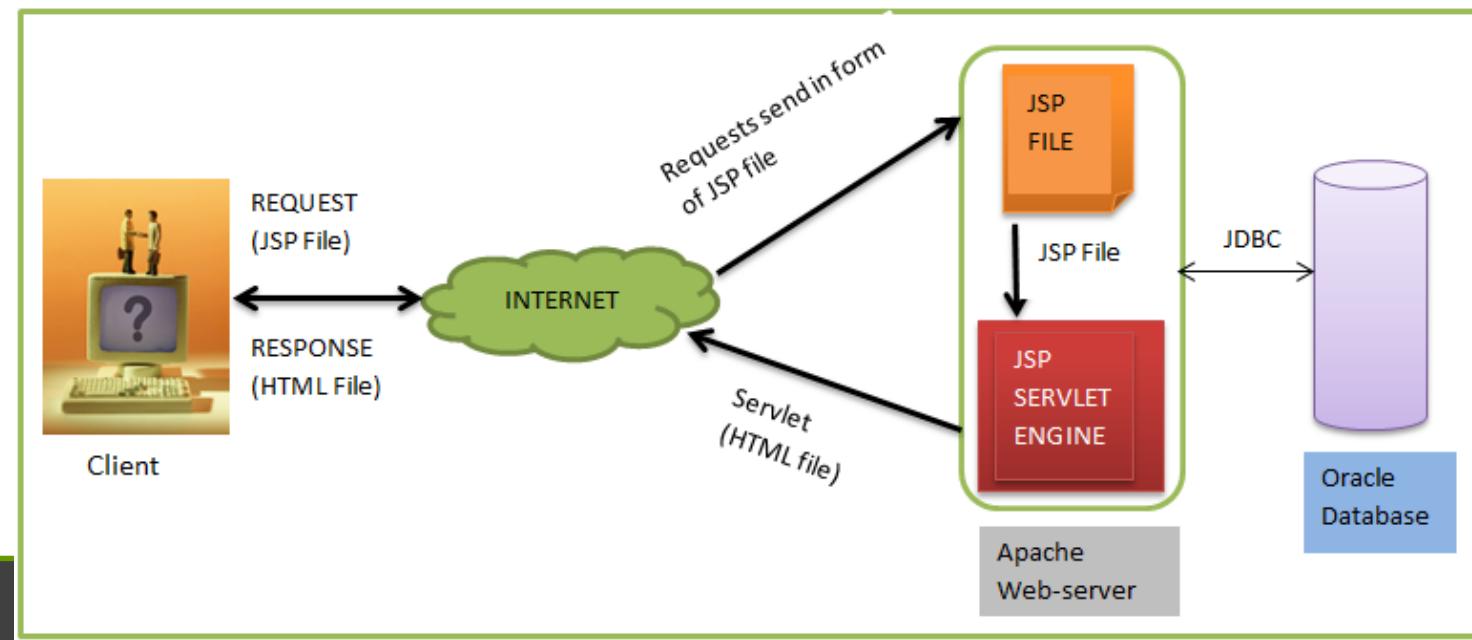
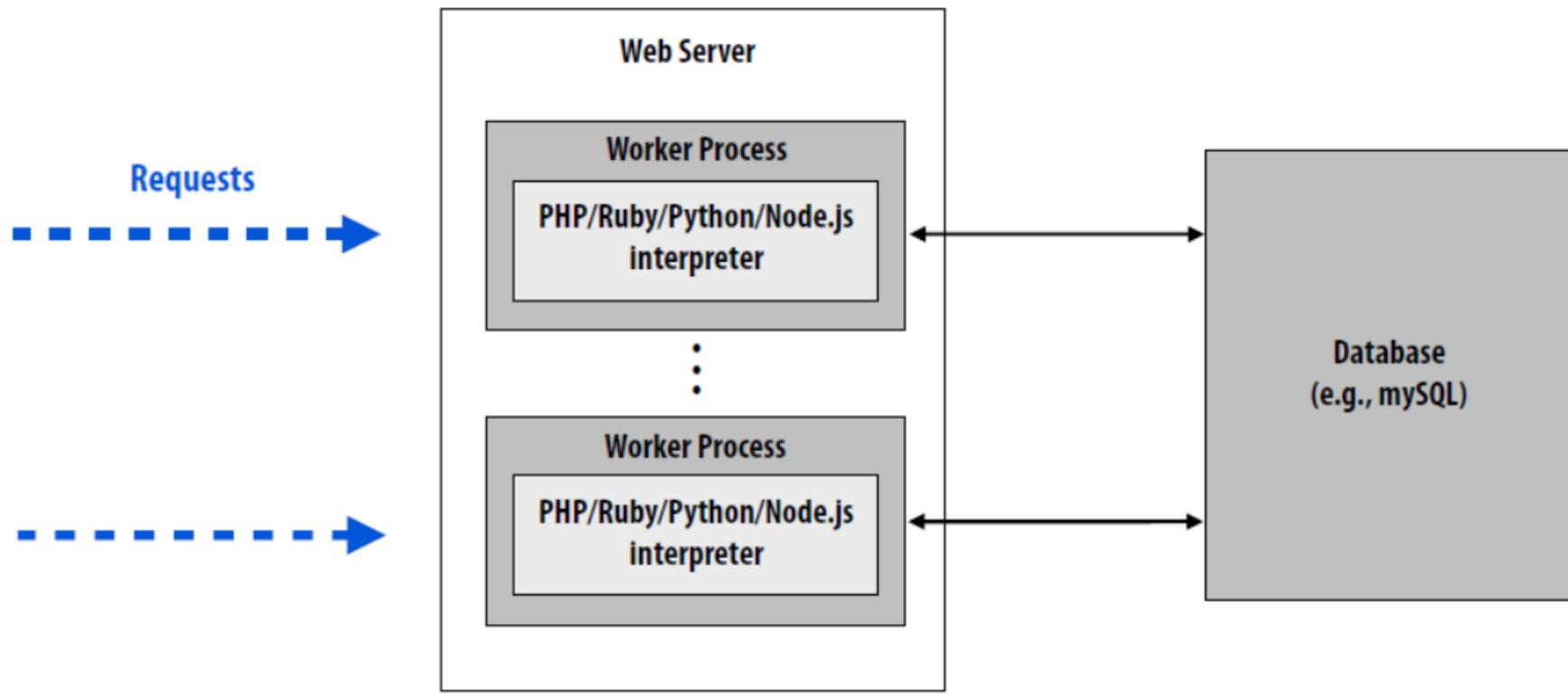
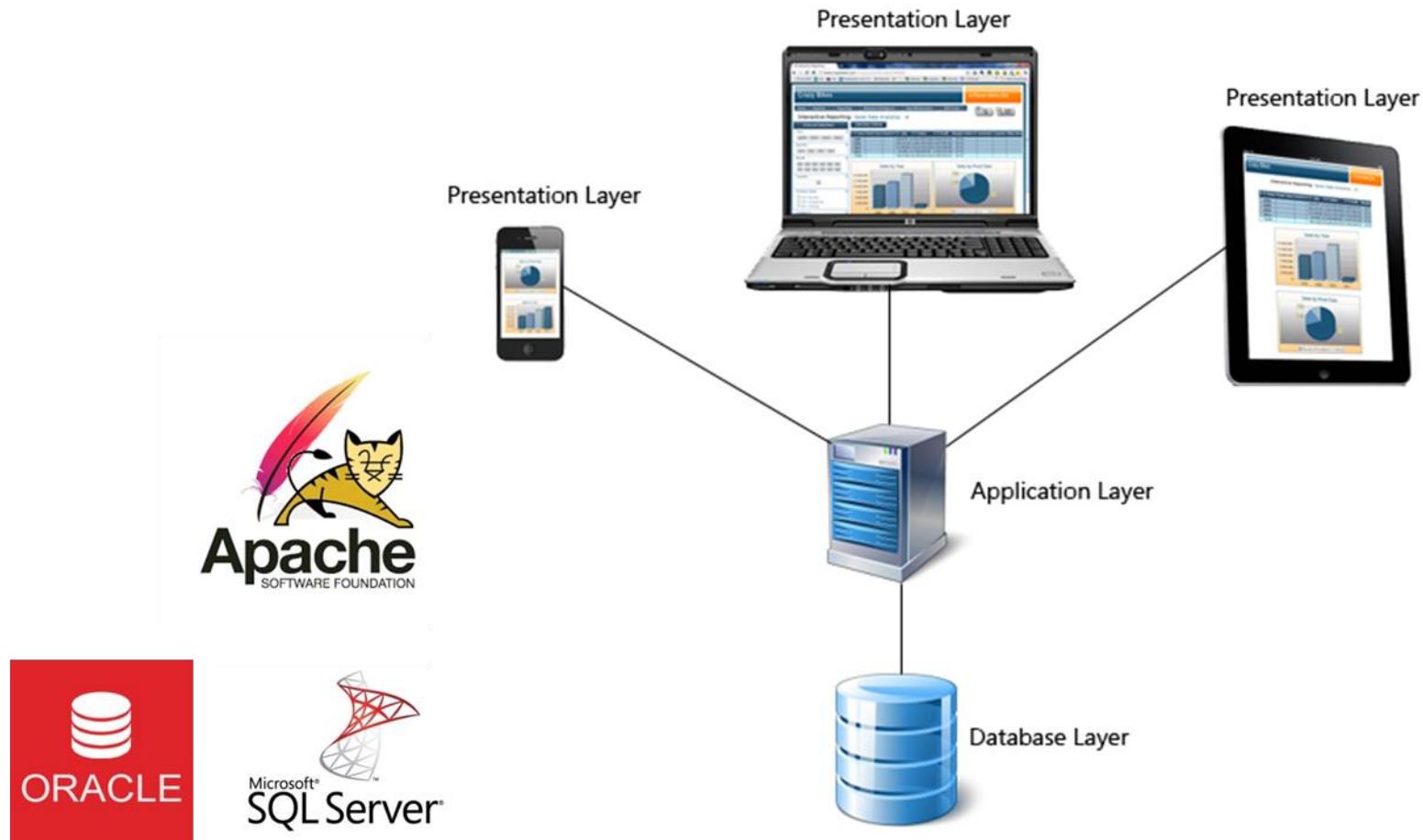


图 1.3 网站使用缓存





❑ Even 3-tier





1.2.4 使用应用服务器集群改善网站的并发处理能力

使用集群是网站解决高并发、海量数据问题的常用手段。当一台服务器的处理能力、存储空间不足时，不要企图去换更强大的服务器，对大型网站而言，不管多么强大的服务器，都满足不了网站持续增长的业务需求。这种情况下，更恰当的做法是增加一台服务器分担原有服务器的访问及存储压力。

对网站架构而言，只要能通过增加一台服务器的方式改善负载压力，就可以以同样的方式持续增加服务器不断改善系统性能，从而实现系统的可伸缩性。应用服务器实现集群是网站可伸缩集群架构设计中较为简单成熟的一种，如图 1.4 所示。



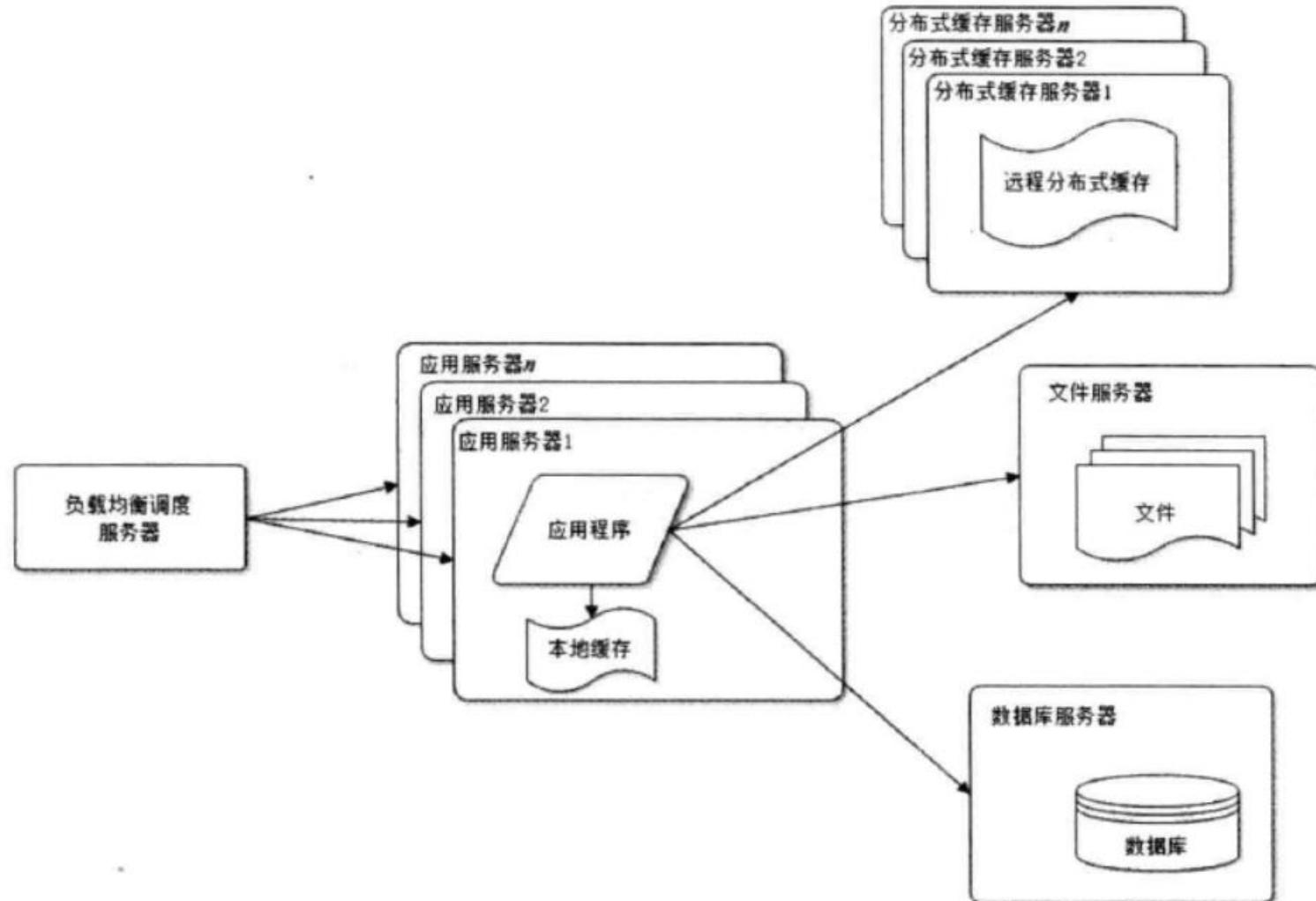
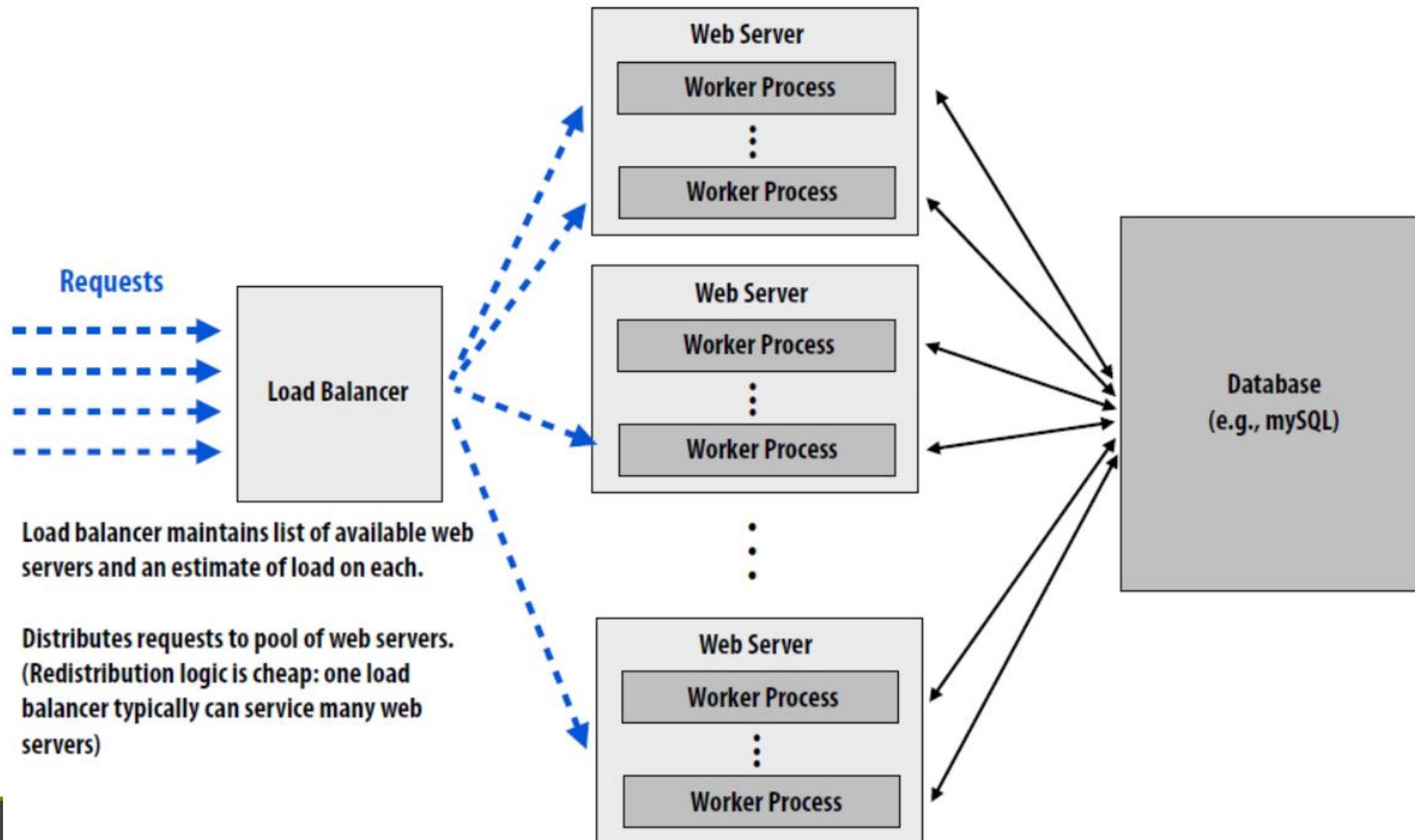


图 1.4 应用服务器集群部署

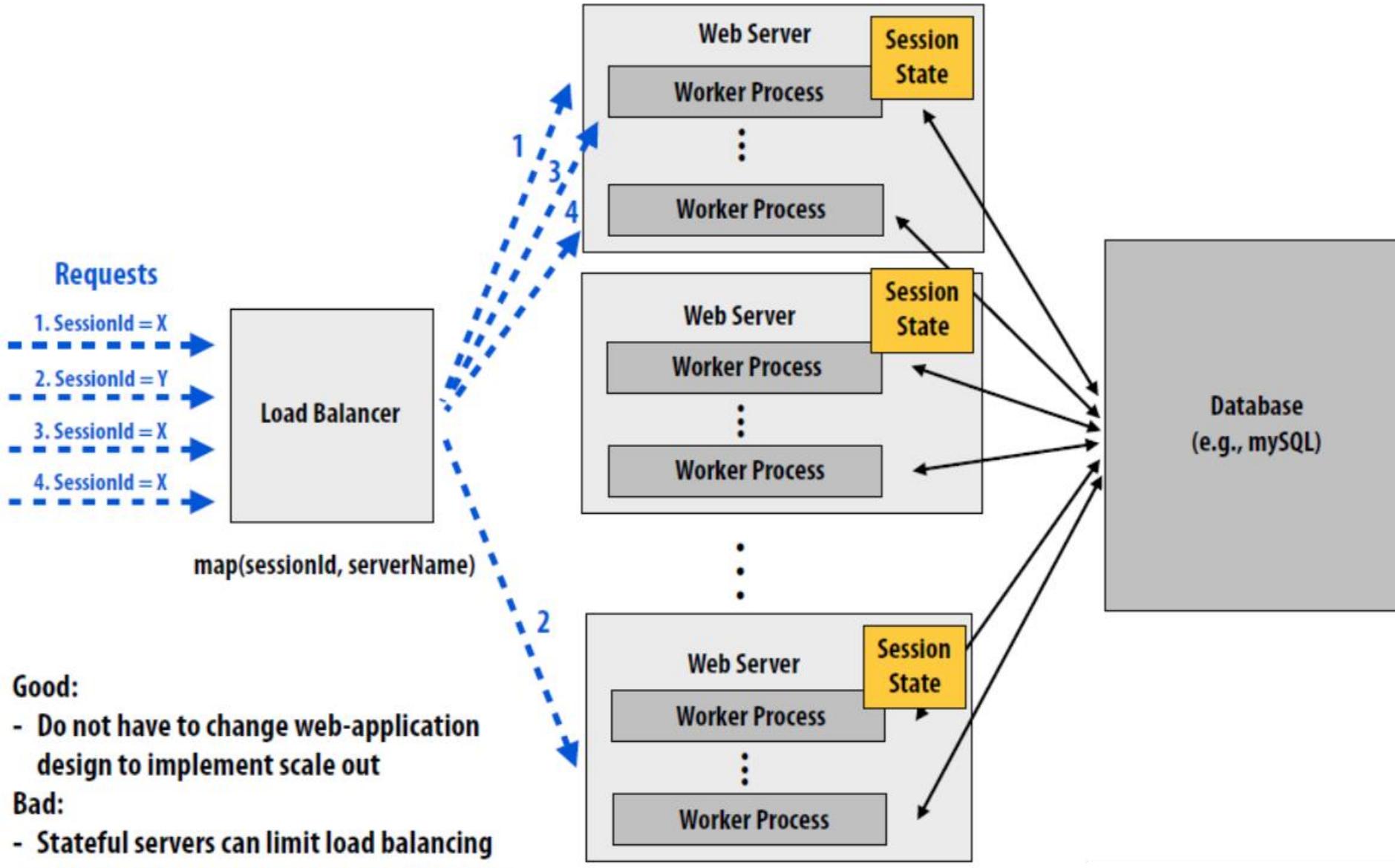
“Scale out” to increase throughput

Use many web servers to meet site’s throughput goals.



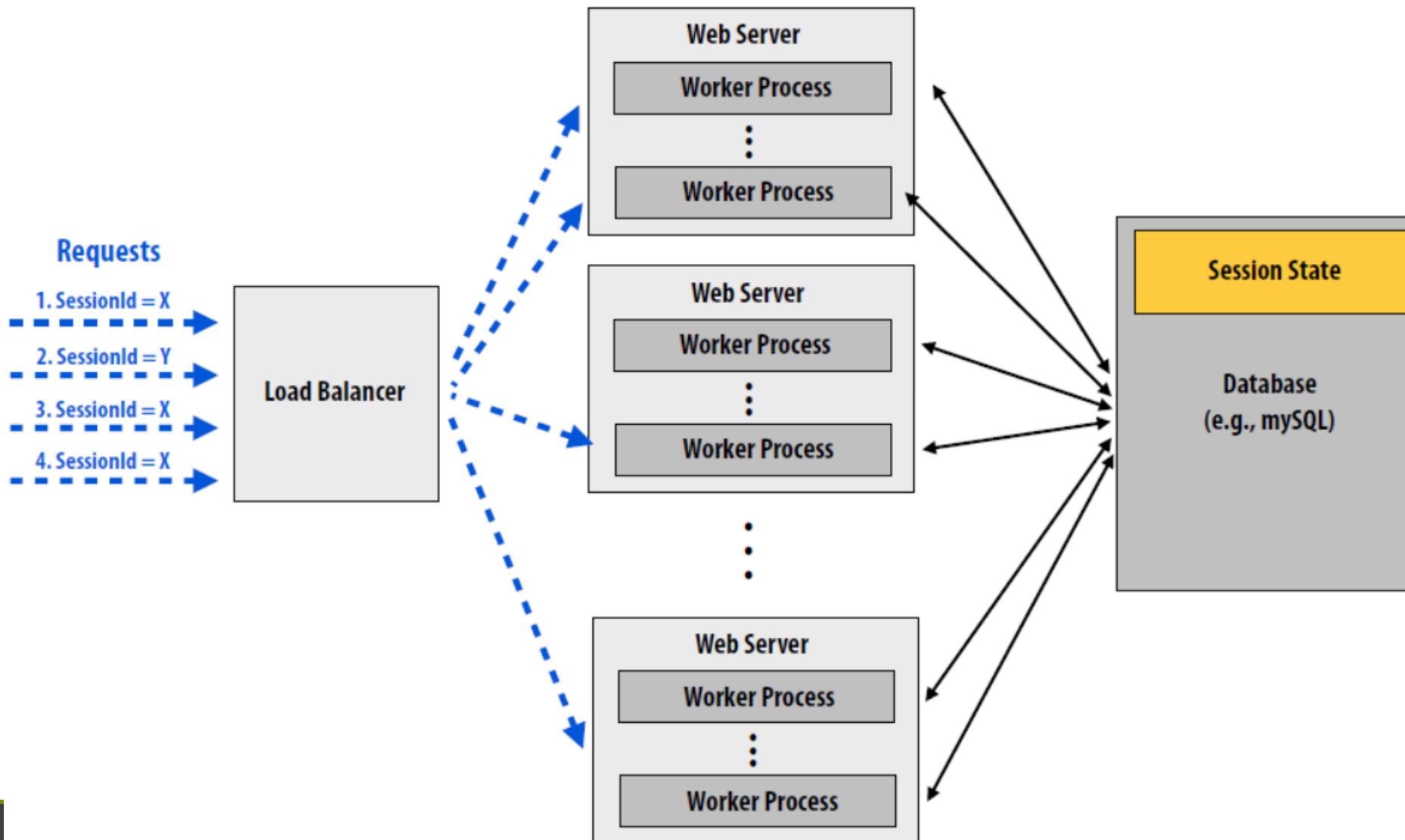
Load balancing with persistence

All requests associated with a session are directed to the same server (aka. session affinity, "sticky sessions")



Desirable: avoid persistent state in web server

Maintain stateless servers, treat sessions as persistent data to be stored in the DB.



1.2.5 数据库读写分离

网站在使用缓存后，使绝大部分数据读操作访问都可以不通过数据库就能完成，但是仍有一部分读操作（缓存访问不命中、缓存过期）和全部的写操作需要访问数据库，在网站的用户达到一定规模后，数据库因为负载压力过高而成为网站的瓶颈。

目前大部分的主流数据库都提供主从热备功能，通过配置两台数据库主从关系，可以将一台数据库服务器的数据更新同步到另一台服务器上。网站利用数据库的这一功能，实现数据库读写分离，从而改善数据库负载压力，如图 1.5 所示。

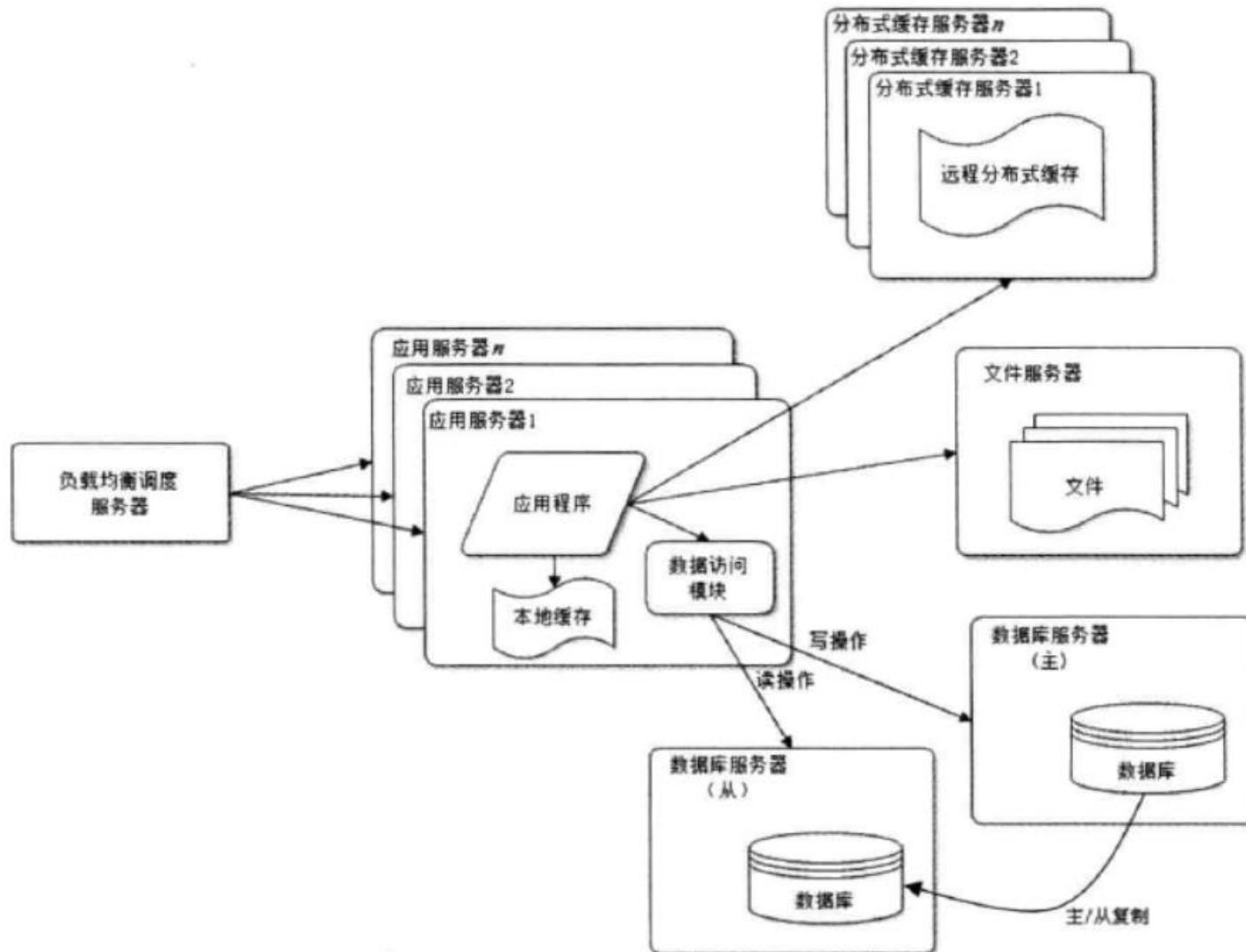


图 1.5 数据库读写分离

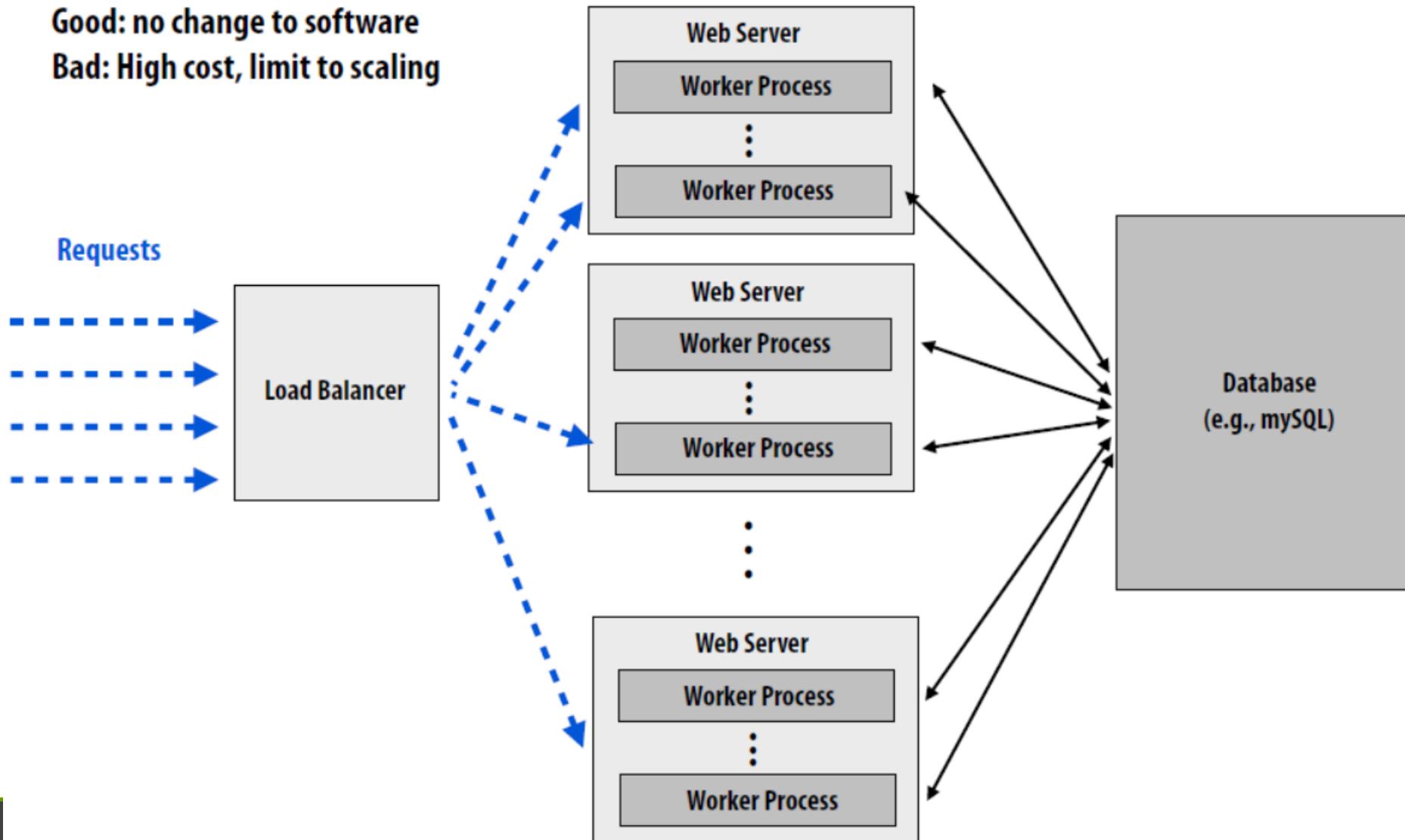
应用服务器在写数据的时候，访问主数据库，主数据库通过主从复制机制将数据更新同步到从数据库，这样当应用服务器读数据的时候，就可以通过从数据库获得数据。

Dealing with database contention

Option 1: "scale up": buy better hardware for database server, buy professional-grade DB that scales
(see database systems course by Prof. Pavlo)

Good: no change to software

Bad: High cost, limit to scaling



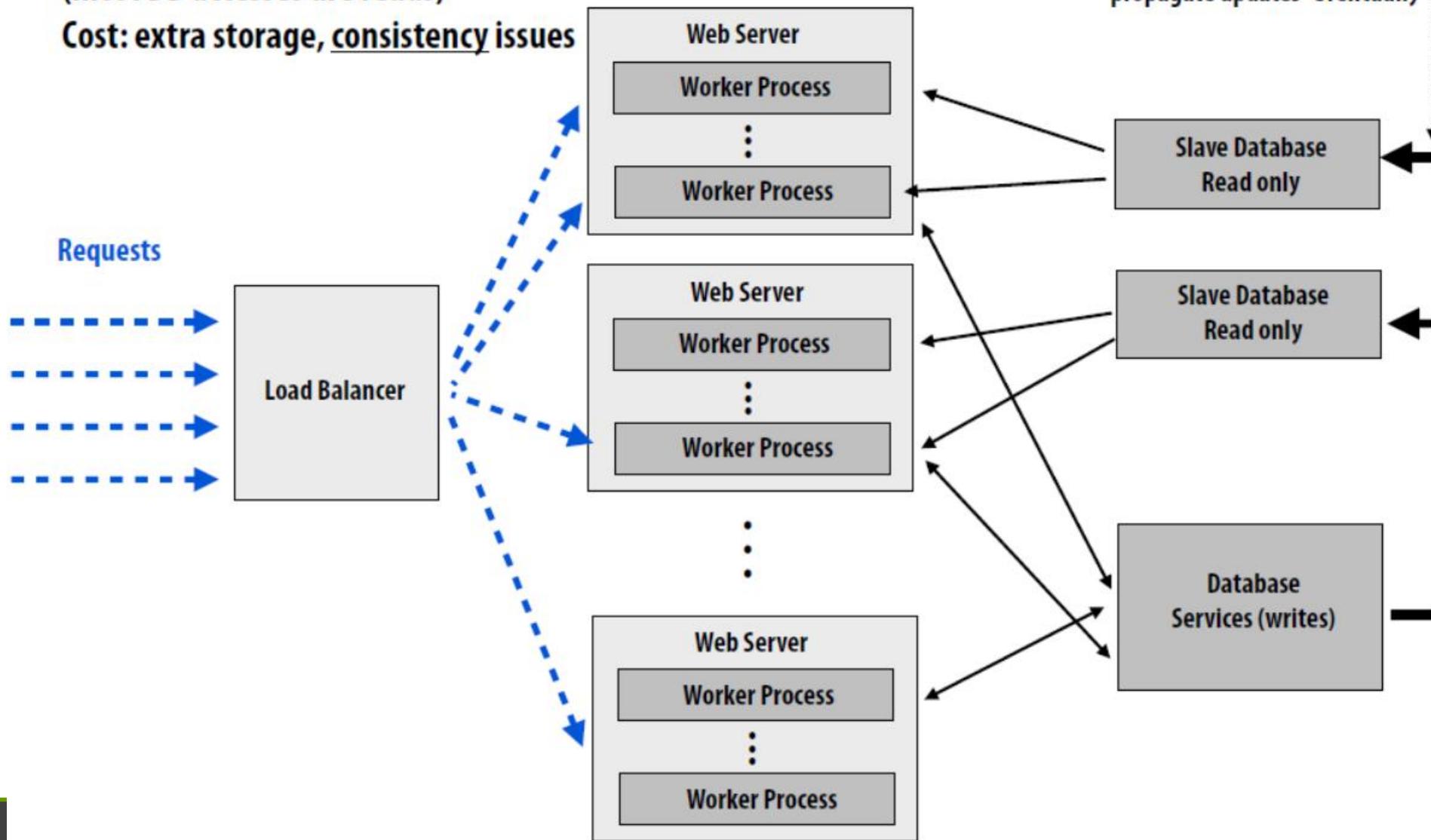
Scaling out a database: replicate

Replicate data and parallelize reads

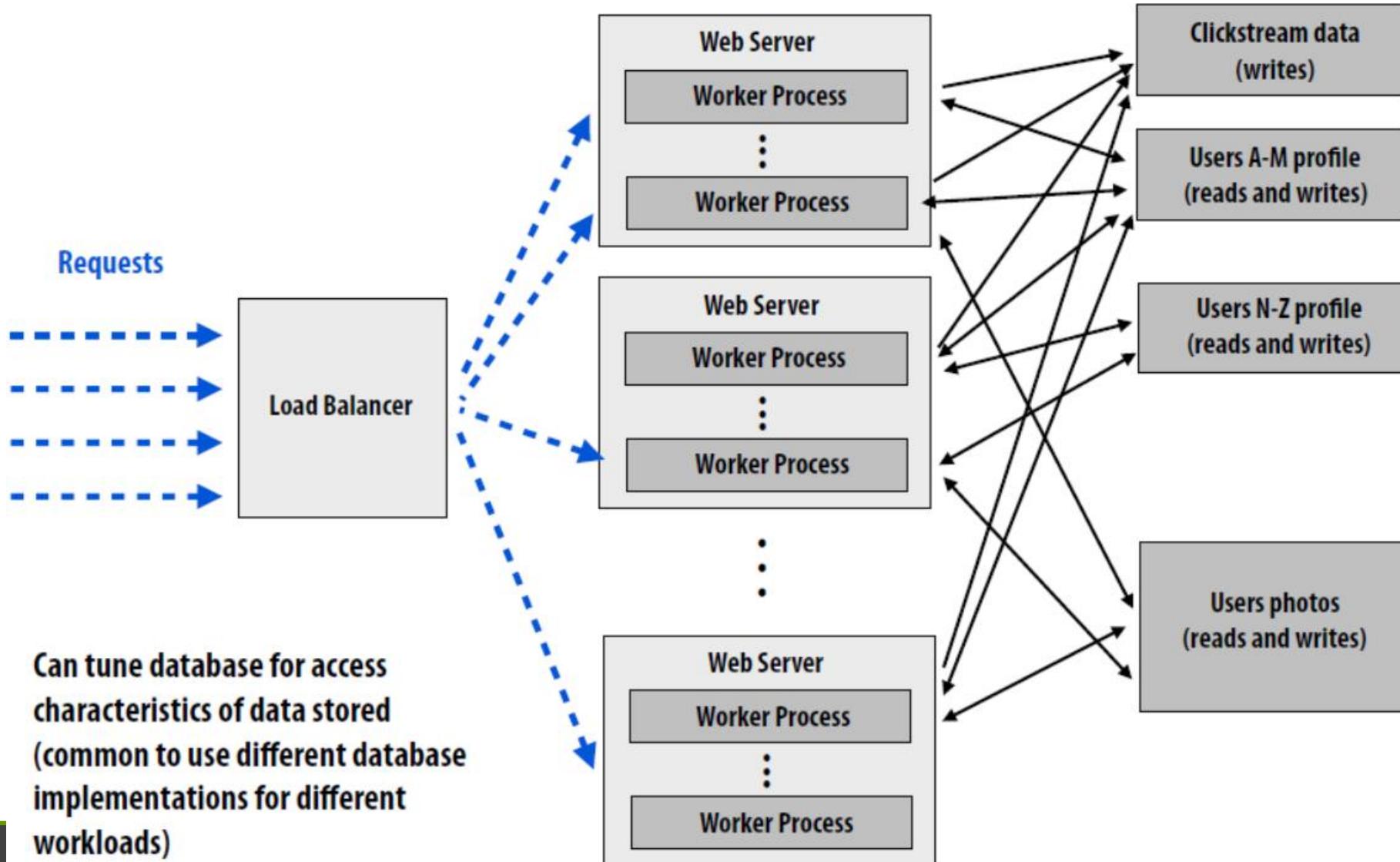
(most DB accesses are reads)

Cost: extra storage, consistency issues

Adopt relaxed memory models:
propagate updates "eventually"

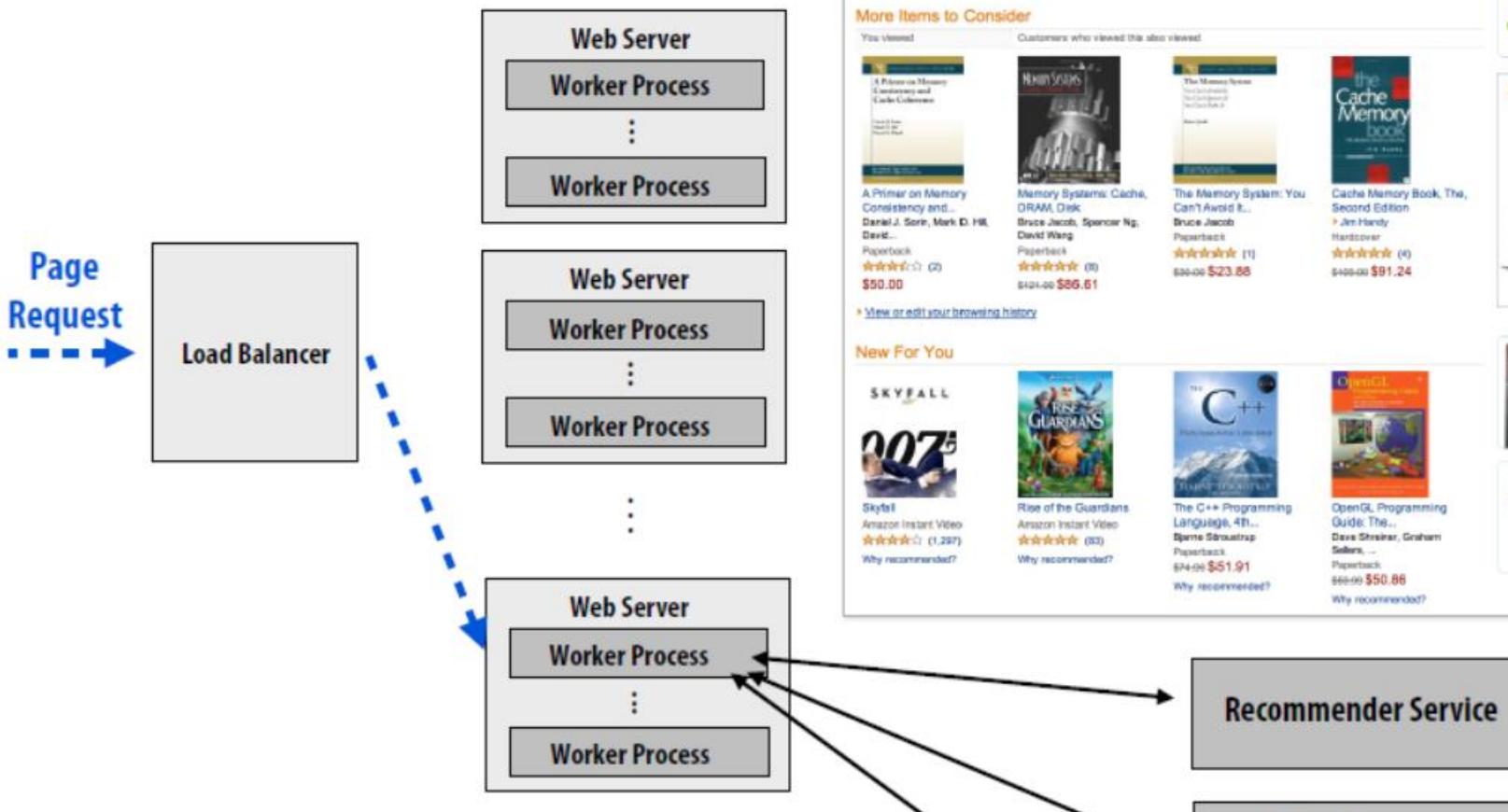


Scaling out a database: partition



Inter-request parallelism

Parallelize generation of a single page



Amount of user traffic is directly correlated to response latency.

See great post:

<http://perspectives.mvdirona.com/2009/10/31/TheCostOfLatency.aspx>

1.2.6 使用反向代理和 CDN 加速网站响应

随着网站业务不断发展，用户规模越来越大，由于中国复杂的网络环境，不同地区的用户访问网站时，速度差别也极大。有研究表明，网站访问延迟和用户流失率正相关，网站访问越慢，用户越容易失去耐心而离开。为了提供更好的用户体验，留住用户，网站需要加速网站访问速度。主要手段有使用 CDN 和反向代理，如图 1.6 所示。

CDN 和反向代理的基本原理都是缓存，区别在于 CDN 部署在网络提供商的机房，使用户在请求网站服务时，可以从距离自己最近的网络提供商机房获取数据；而反向代理则部署在网站的中心机房，当用户请求到达中心机房后，首先访问的服务器是反向代理服务器，如果反向代理服务器中缓存着用户请求的资源，就将其直接返回给用户。



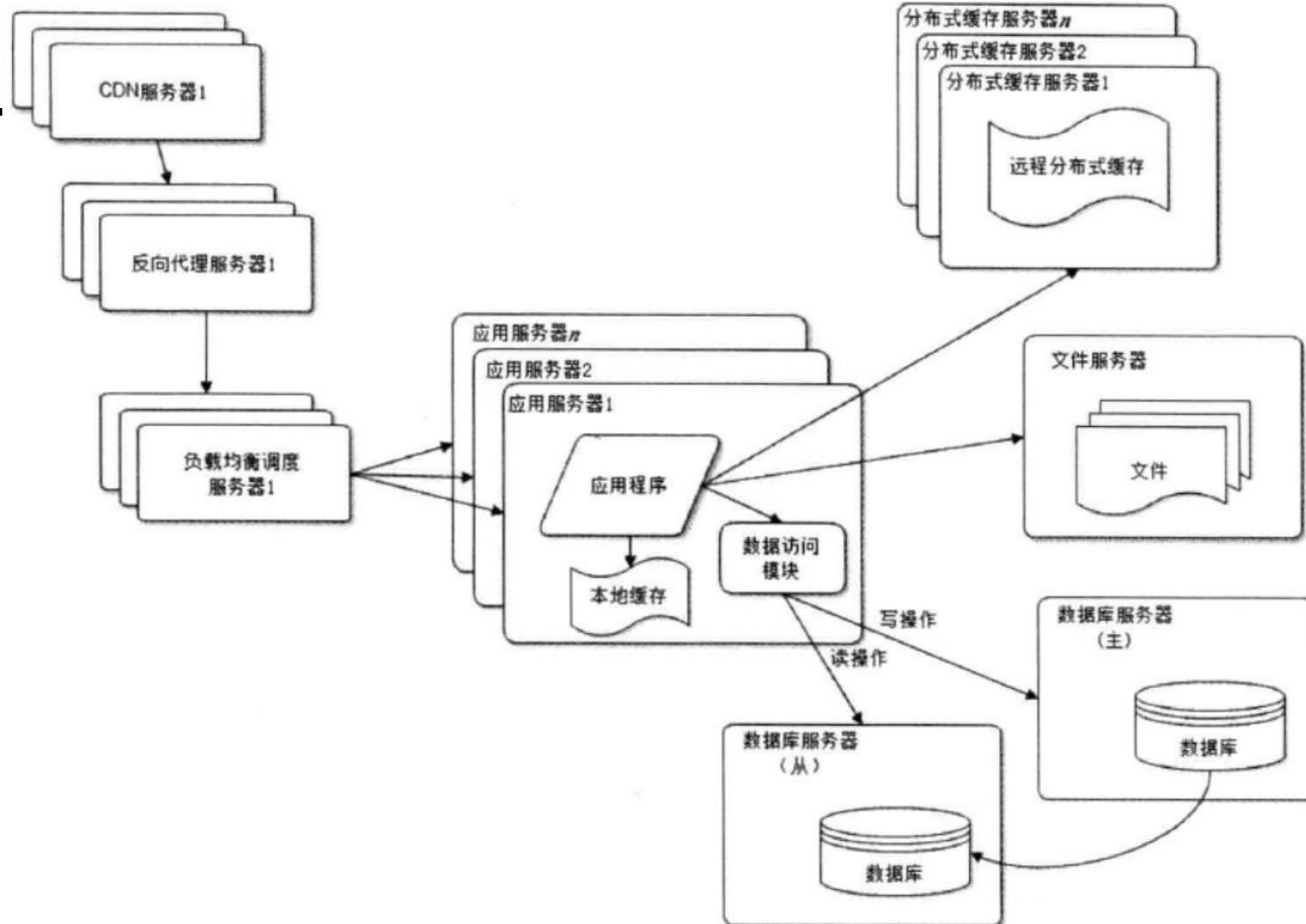
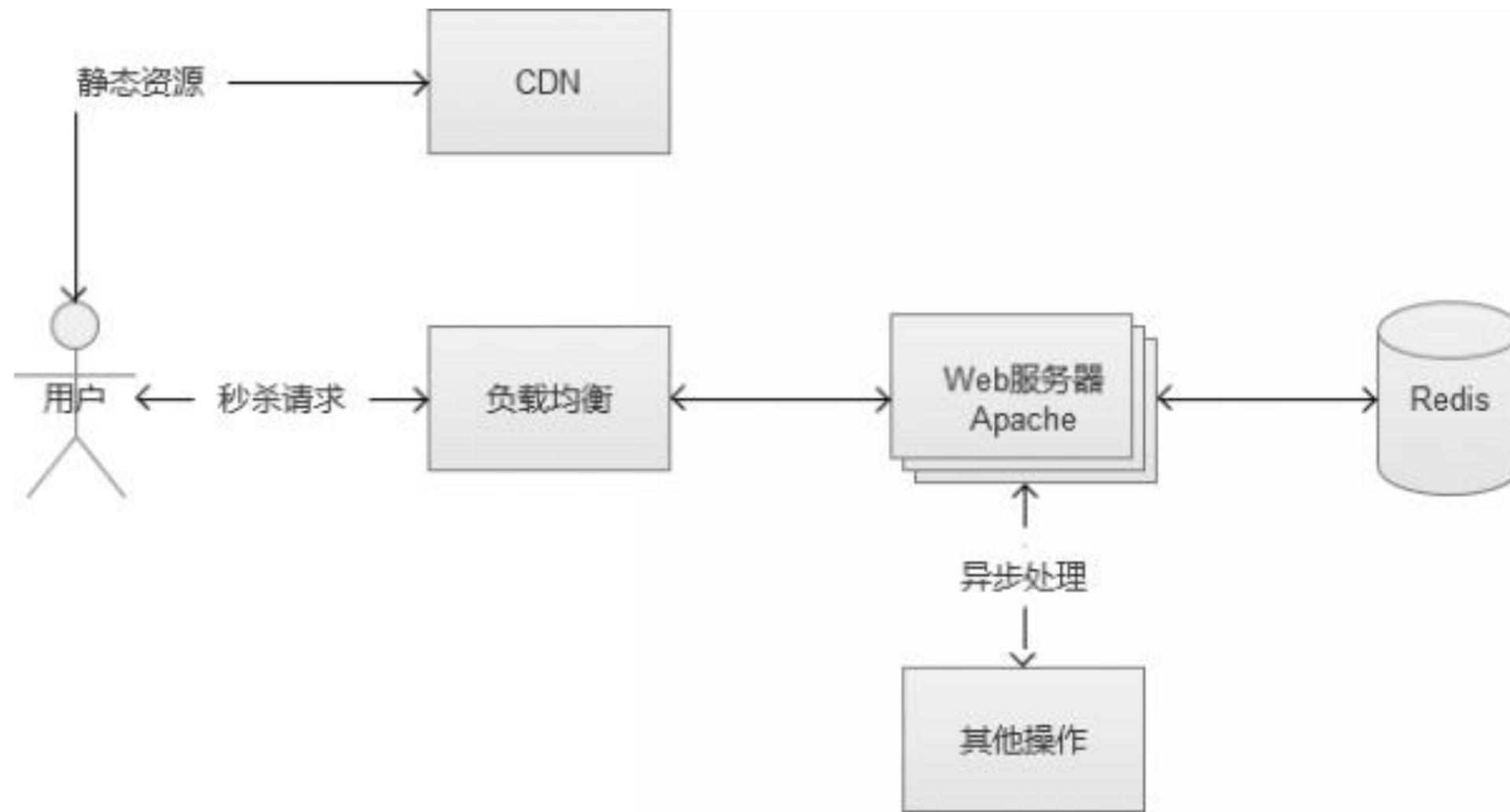
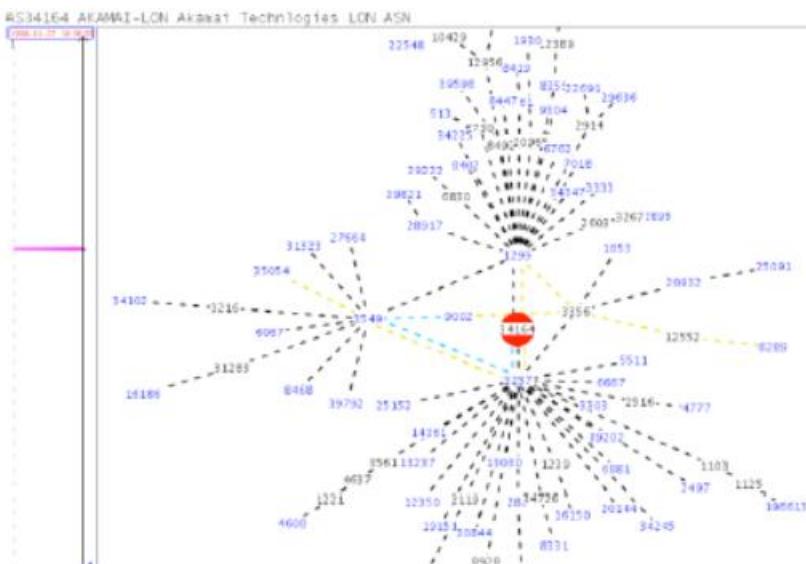
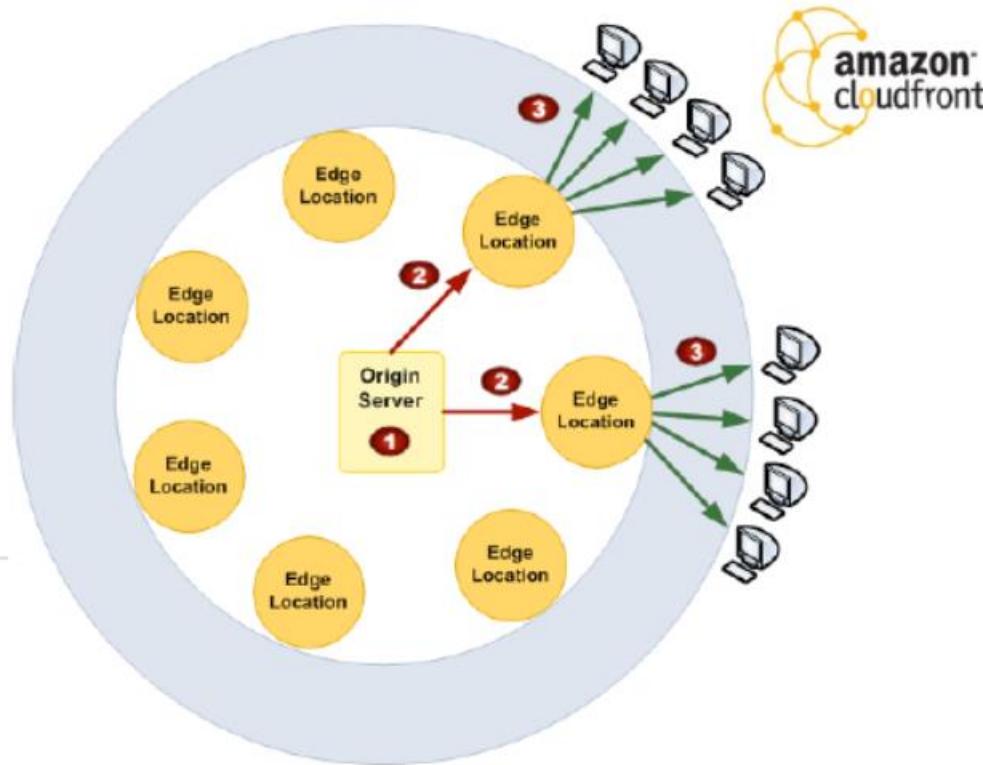


图 1.6 网站使用反向代理和 CDN 加速访问



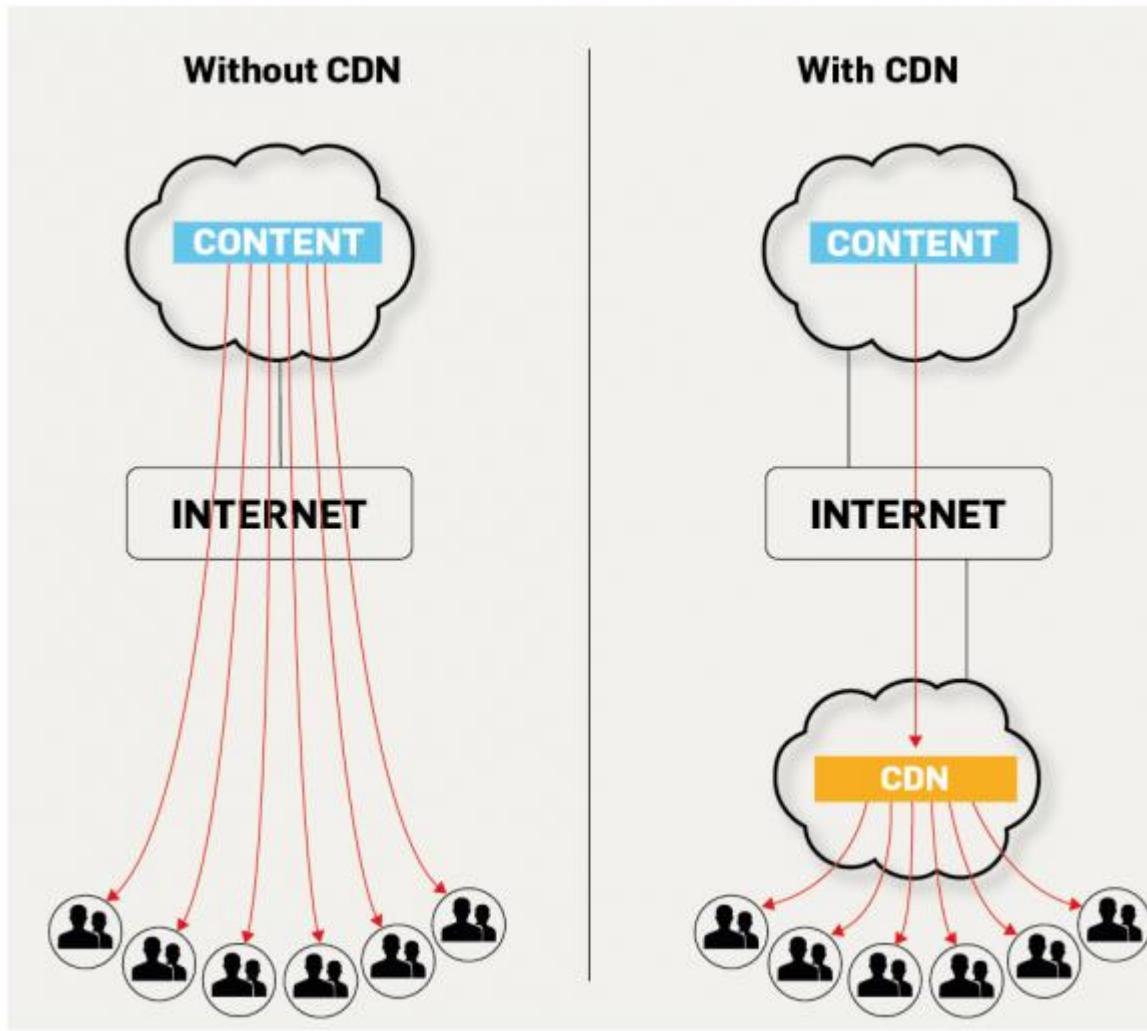
Caching using content distribution networks (CDNs)

- Serving large media assets can be expensive to serve (high bandwidth costs, tie up web servers)
 - E.g., images, streaming video
- Physical locality is important
 - Higher bandwidth
 - Lower latency



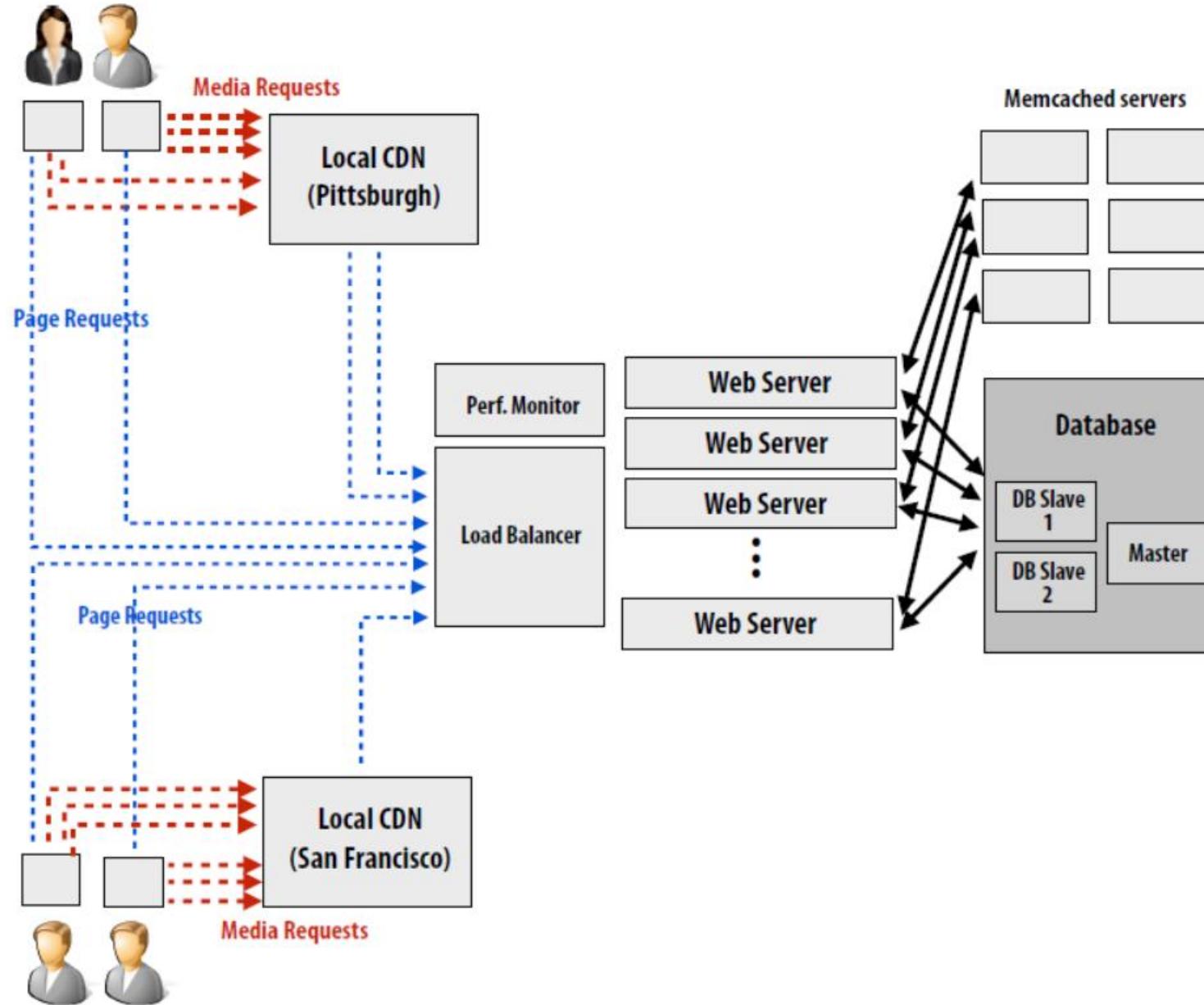
London Content Distribution Network

Source: http://www.telco2.net/blog/2008/11/amazon_cloudfront_yet_more_tra.html





CDN integration



Chapter 5: Content Delivery Network Companies

The industry has grown significantly in the recent years, as the importance of a CDN has also grown. The Internet is always evolving and so is the need for faster websites and better user experience. Nowadays, there are more and more CDN providers, but here is a quick list of the top 12:

1. Akamai

<https://www.globalsign.com/content-delivery-network-explained/>

Akamai is known globally to speed up networks and connections. They have a strong infrastructure to speed up sites and make things better for online users. Of course, this company works with big brands and large businesses in general. It's not commonly used by bloggers, but by customers with a huge page traffic and very well established brands.



2. MaxCDN

MaxCDN was founded in 2009 and is based in Los Angeles, California. MaxCDN has fast servers and a big community of users, and also, webmasters from around the World. They have one of the fastest response time and flexible pricing models available. Their CDN is almost always used for WordPress, Joomla, Drupal, OpenCart, PrestaShop, and all the other applications and it utilizes Anycast stateless routing for one-to-nearest content delivery over multiple 10 Gbit/s connections. It's mainly used for websites, blogs and gaming platforms.



3. Incapsula

Incapsula Inc. is a Cloud-based application delivery platform. It uses a global content delivery network to provide website security, DDoS protection, load balancing and failover services to clients. It's stationed in Redwood Shores, CA. They offer CDNs paired with Web Security combined with smart balancer technology that handle the traffic between servers. The company works with popular sites like Moz, Wix, SIEMENS and others.



4. Rackspace

Founded back in 1998, Rackspace Inc. is a cloud computing company based in Windcrest, Texas. Rackspace is unique for its cloud file storage and their "pay as you go" business model. It's a secure environment to host a large file and static websites. It's a well protected infrastructure, with almost two decades of experience, customer trust and brand awareness.



5. Cloudflare

CDN

- 京东从2009年开始自建CDN服务
- 承担90%互联网流量
- 需要解决一些**性能和结构**上的问题

负载均衡

- 2013年双11时，京东主要负载均衡业务使用的是商用产品
- 随着业务发展，商业负载均衡产品暴露出一些突出问题：**负载过高、宕机增加、不易扩容**

基础网络

- 2013年双11时，数据中心**TOR上联**带宽拥塞的问题愈发突显
- 大量的跨机房服务调用和数据传输需求，对POD间、DC间的**专线**带宽和可靠性要求越来越高
- 机架不足

CDN、负载均衡、基础网络面对的主要问题

CDN 面对的问题

2009年，运维部元老Gavin在对CDN几乎零经验的情况下，通过在网上查找资料自学、大量实验和环境测试，搭建了基于Squid实现的第一版京东CDN服务。出于好奇我特别找到Gavin了解当时决定自建CDN的原因，原来是因为当时使用的商用CDN服务出现了几次严重的故障，促使Gavin决定自己来做，看来自己动手的基因真的是深入互联网公司的骨髓。

CDN 提升

2013年上半年京东 CDN 全部改造为 HAProxy + ATS 的结构。

接下来，首先实现了将 HAProxy 的单机性能提升一倍，在此基础上再将 HAProxy 的多个 1000M 网卡绑定方式升级为单个 10G 网卡接入，这样在除了网卡之外硬件配置不变的情况下就显著地提升了单台 HAProxy 的服务能力，这是商用产品很难达到的。

当然，设备数量增加、功耗增加、结构复杂度增加、管理维护成本增加也是显而易见的，但相对收益来讲这些代价几乎可以忽略不计。

在过去的一年里，软件负载均衡单机性能提升一倍，集群容量提高 10 倍，替代了 90% 的商用负载均衡设备，将接近 90% 的业务逐步迁移到软负载产品上，软负载集群规模扩大了 10 倍左右，服务中断时间大幅度下降。在开发和使用过程中，针对消除缺陷、性能提升和功能提升累计开发了三十多个补丁，增加了内核监控模块实现对 VIP 的 Bps/pps/qps 流量监控和面向业务颗粒的监控，解决了 HAProxy 多 CPU 负载不均、Keepalived 特定版本因 Gratuitous-ARP 实现缺陷导致 VIP 主备切换失效等相对复杂的问题，并在 2014 年 618 之前实现了 HTTP Anti-flood 的一期功能。

□ CDN服务器是一种新型的网络服务器构建方式。

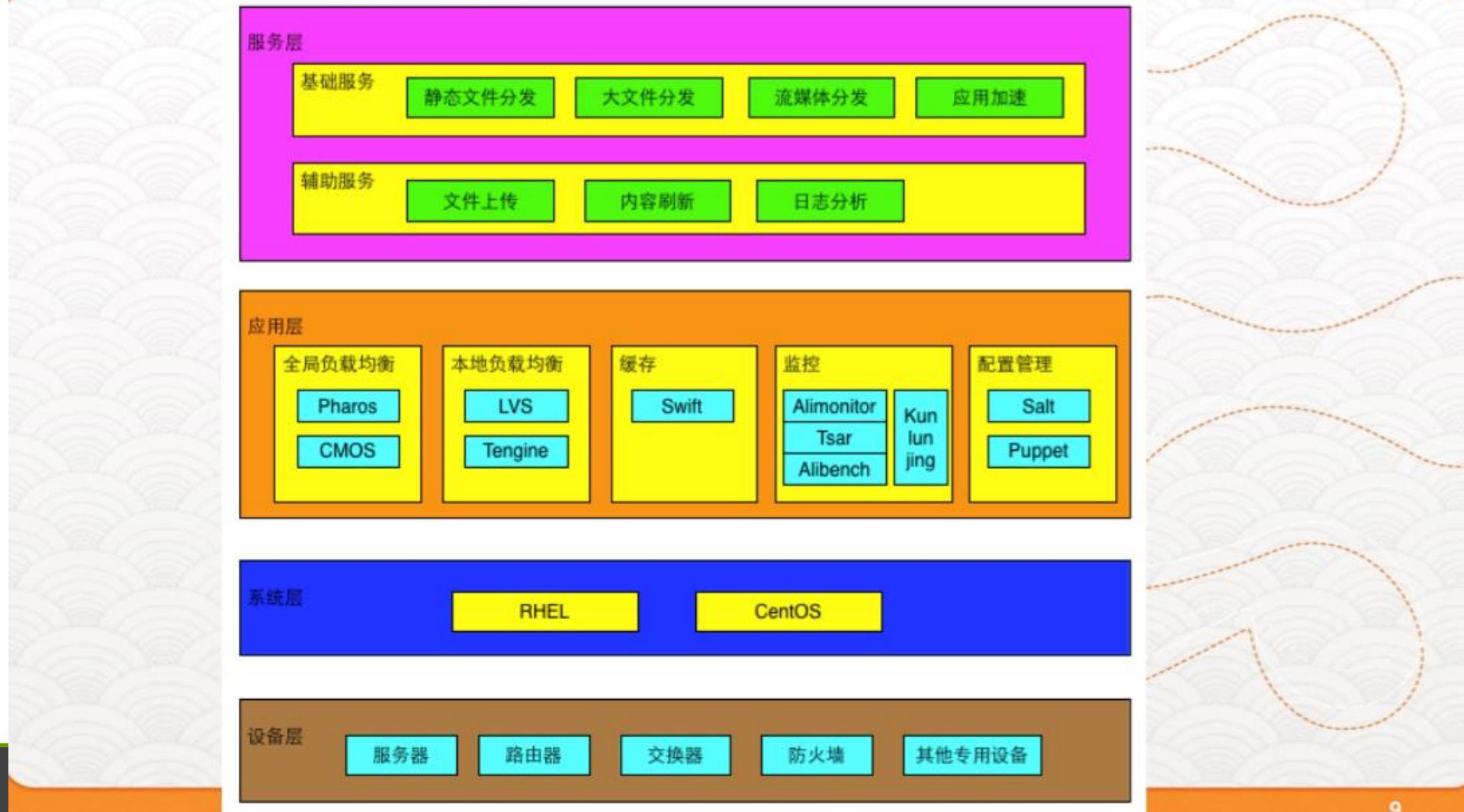
- 如今我们搭建的网站项目面向的用户群比较广泛，有国内、国外的客户群，尤其是在国内用户群中我们还有电信、移动、联通等等不同的线路区别。
- 一般我们购买的虚拟主机、VPS服务器只有一个IP地址，肯定没有哪个商家确保100%线路的稳定和速度，为了优化原有的网络架构服务，因此CDN服务器孕育而生。

□ 搭建自己CDN服务器，从而实现网站加速，对企业来说是很有必要的。

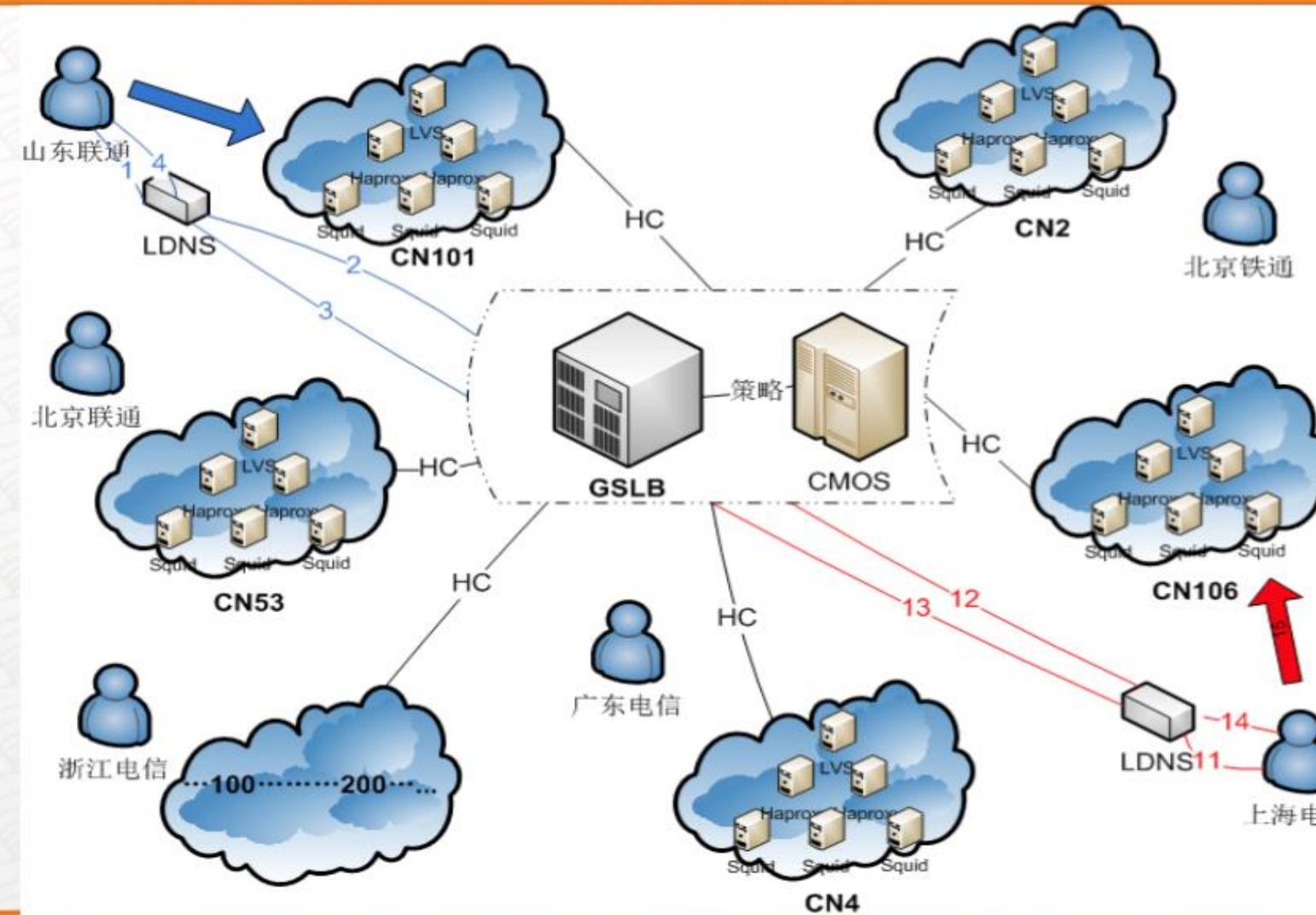
- 一般的企业在建站的时候会有考虑使用CDN加速，有免费和付费的，无论是免费或者付费的，原理都是差不多的，通过多节点让访客地区用户接近节点，提高访问速度。

阿里云CDN技术演进

阿里CDN组件分层

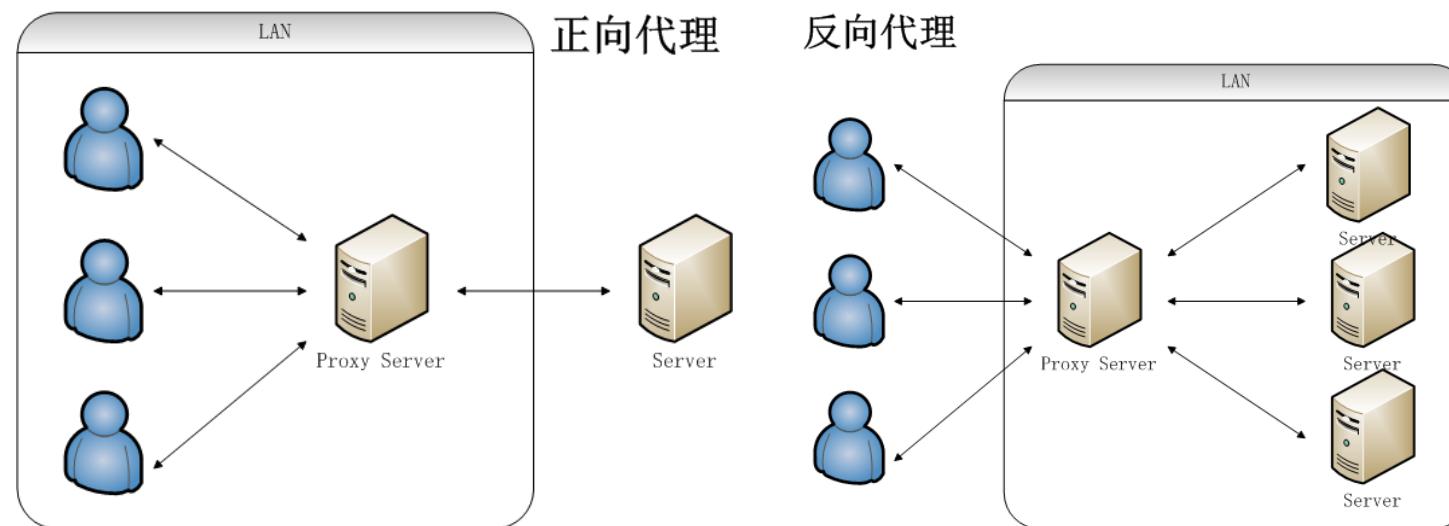


阿里CDN的大脑：全局流量调度

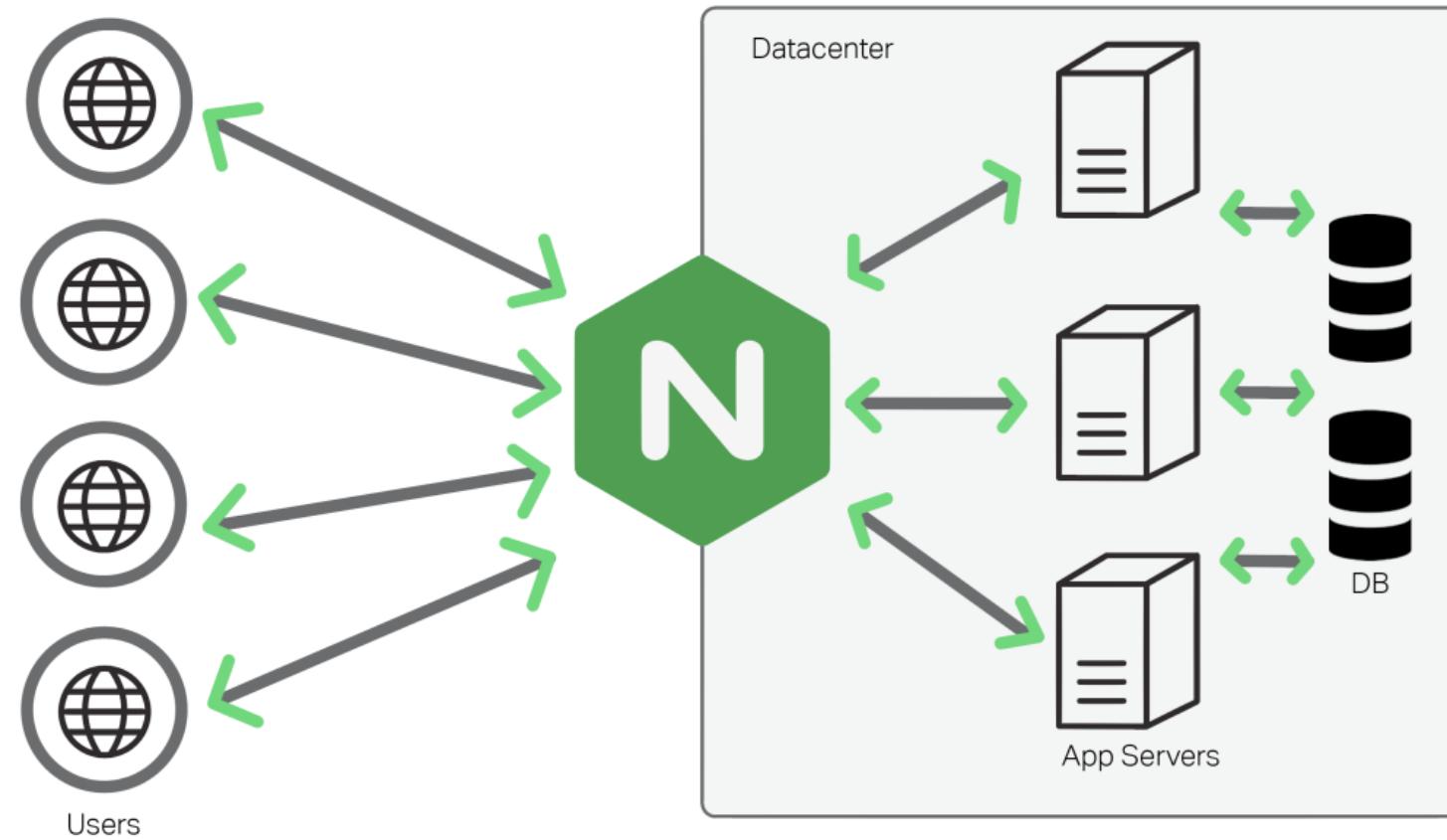


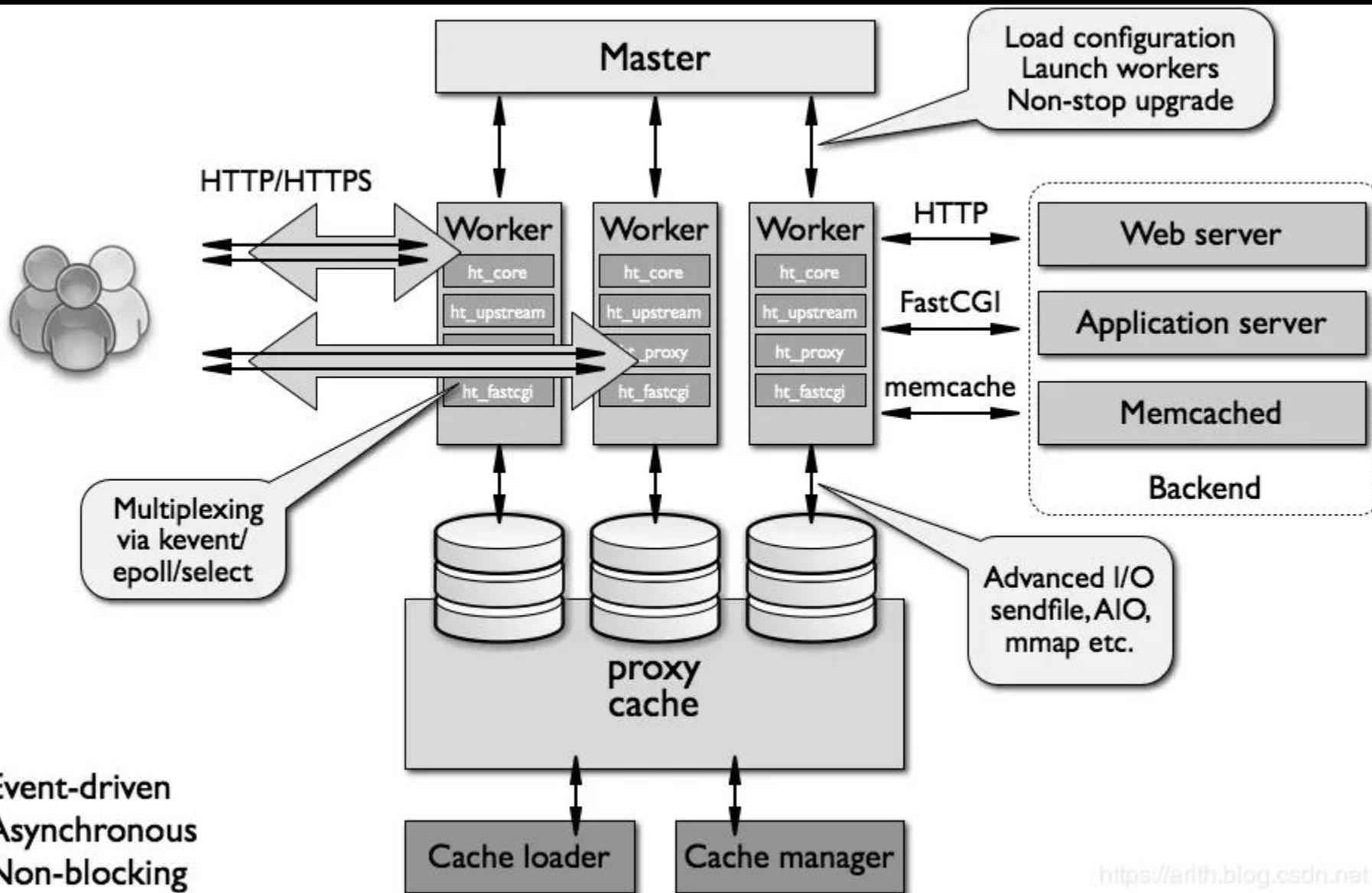
反向代理 (Reverse Proxy)

- 反向代理 (Reverse Proxy) 方式是指以代理服务器来接受Internet上的连接请求，然后将请求转发给内部网络上的服务器，并将从服务器上得到的结果返回给Internet上请求连接的客户端，此时代理服务器对外就表现为一个服务器。



□ Nginx 是很流行的反向代理（也具有Web Server 的功能）

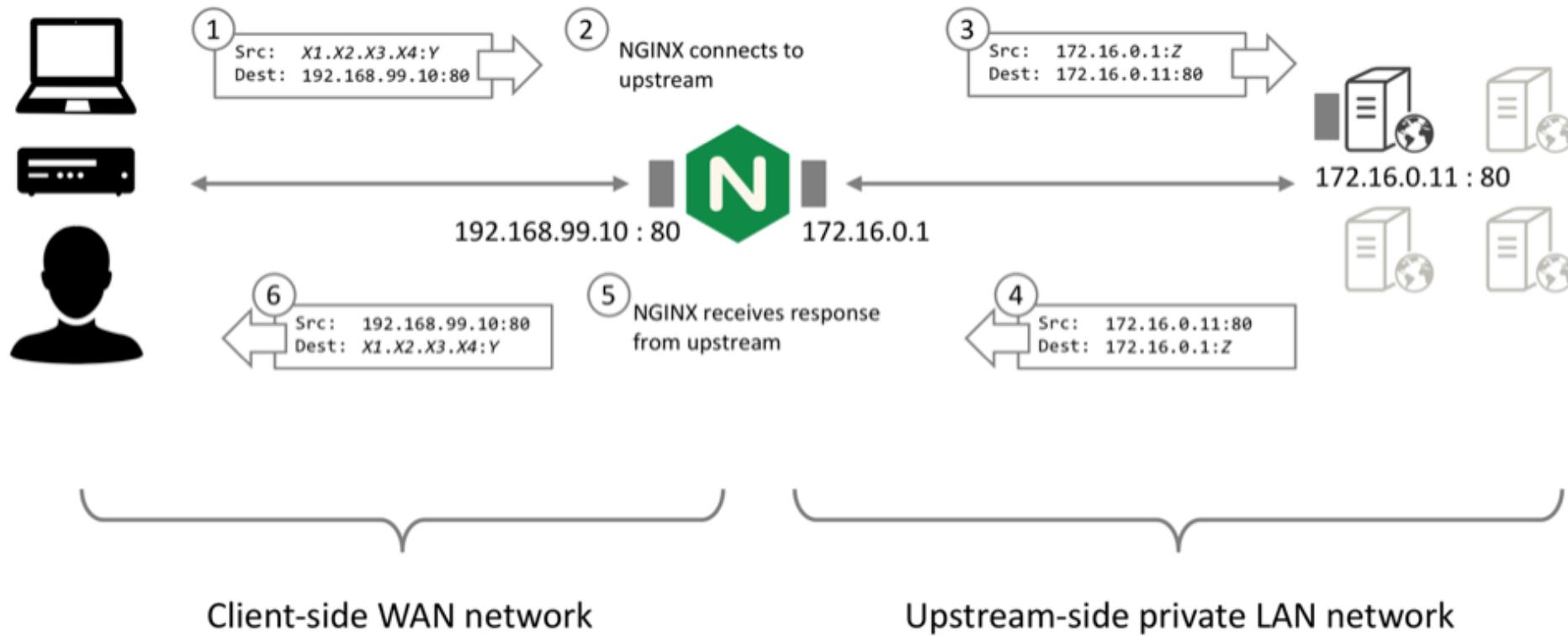




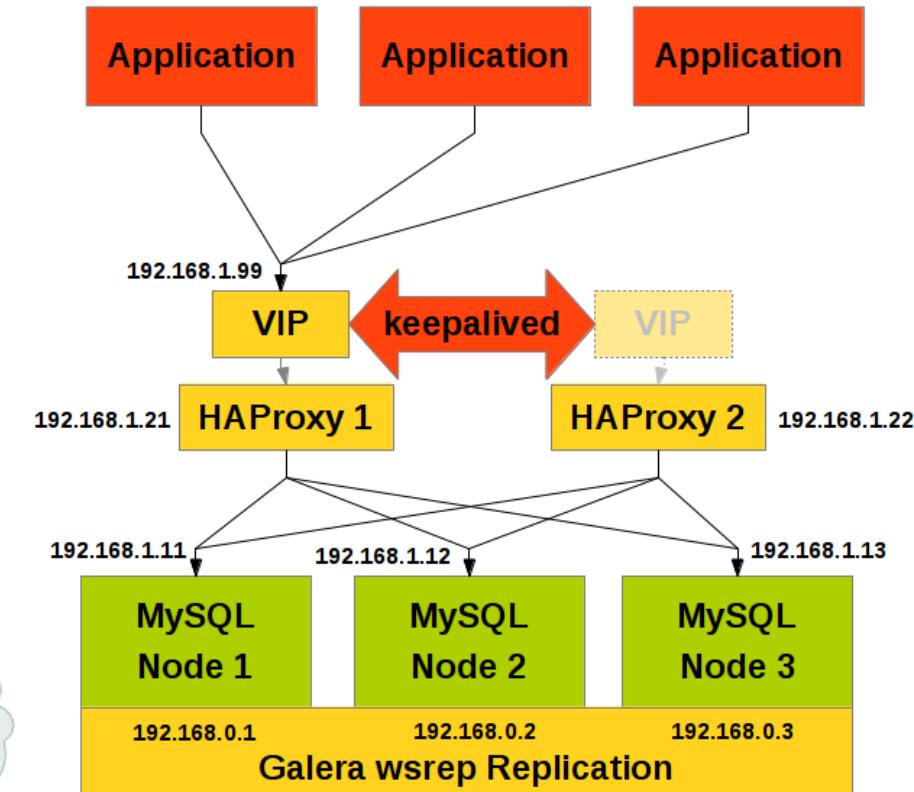
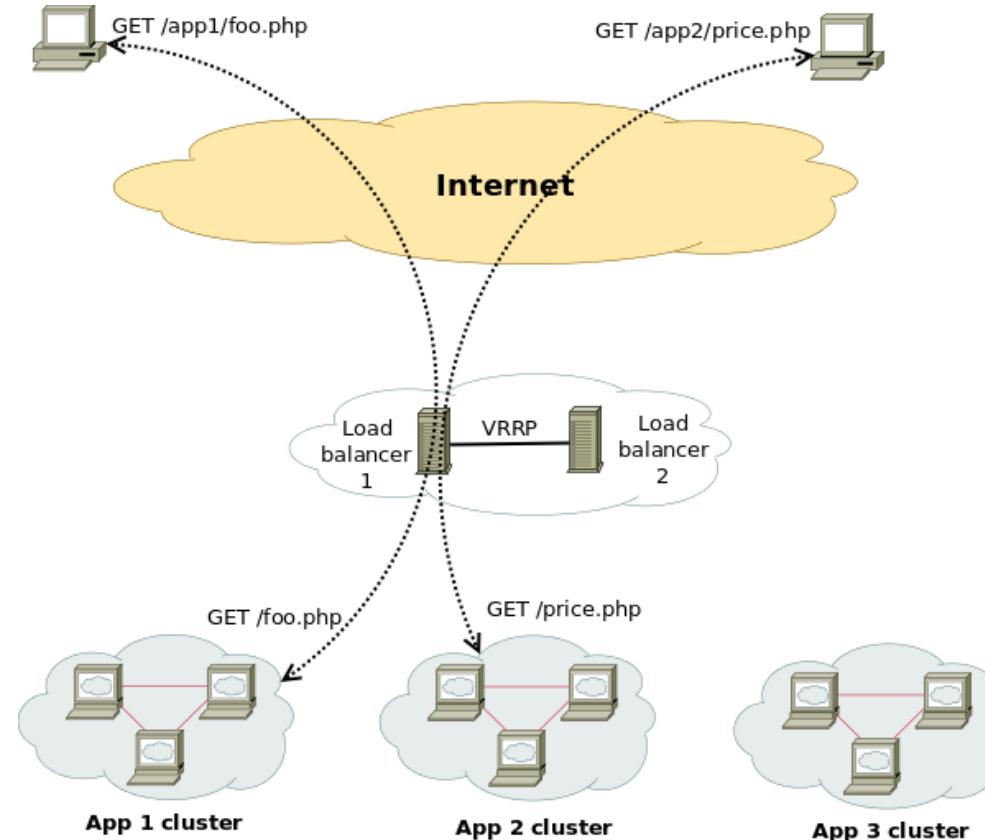
- Event-driven
- Asynchronous
- Non-blocking

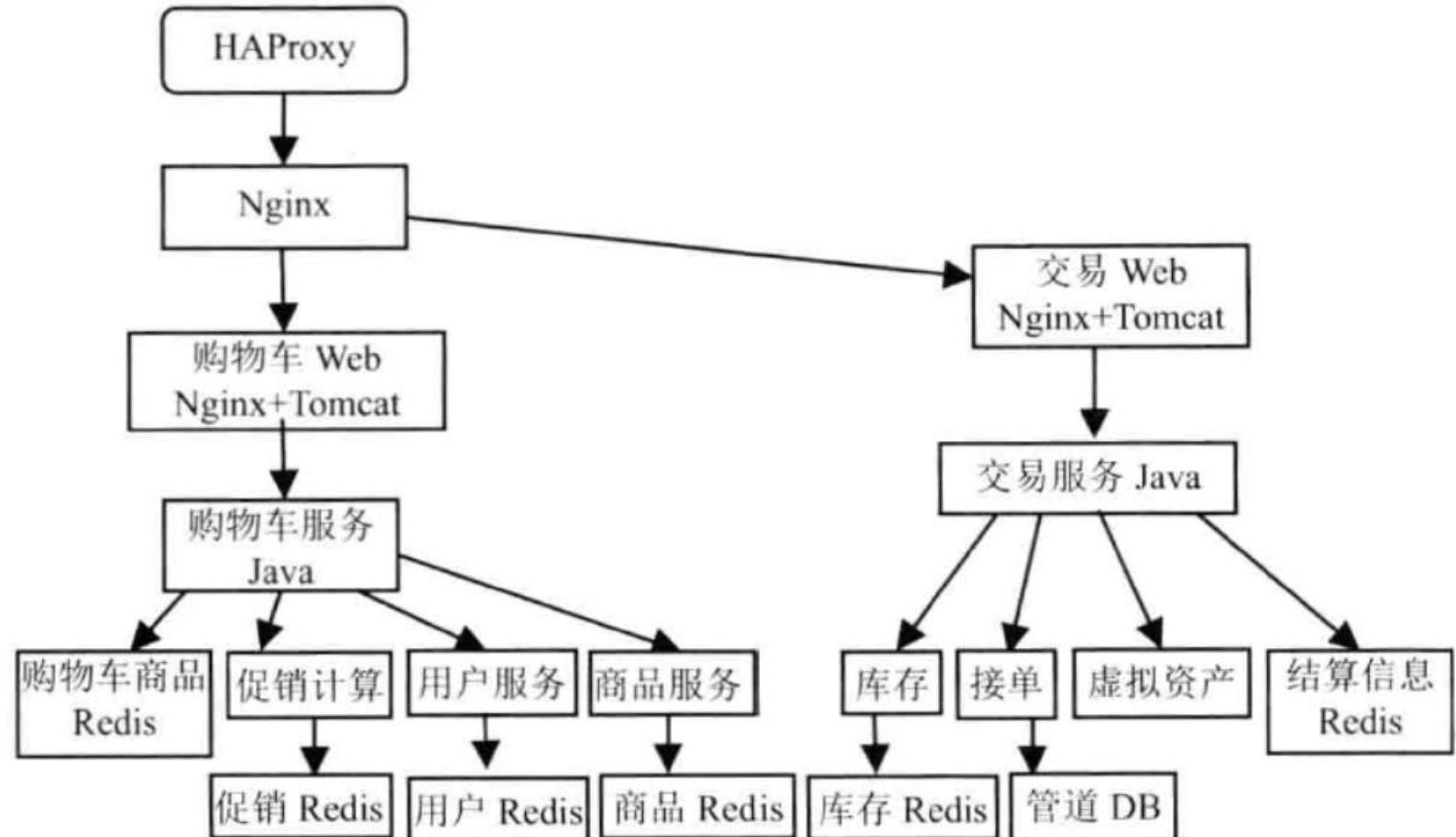
Remote clients access
<http://192.168.99.10/>

Upstream servers on
172.16.0.11-14



HAProxy





京东交易系统的系统架构图

1.2.7 使用分布式文件系统和分布式数据库系统

任何强大的单一服务器都满足不了大型网站持续增长的业务需求。数据库经过读写分离后，从一台服务器拆分成两台服务器，但是随着网站业务的发展依然不能满足需求，这时需要使用分布式数据库。文件系统也是一样，需要使用分布式文件系统，如图 1.7 所示。

分布式数据库是网站数据库拆分的最后手段，只有在单表数据规模非常庞大的时候才使用。不到不得已时，网站更常用的数据库拆分手段是业务分库，将不同业务的数据部署在不同的物理服务器上。



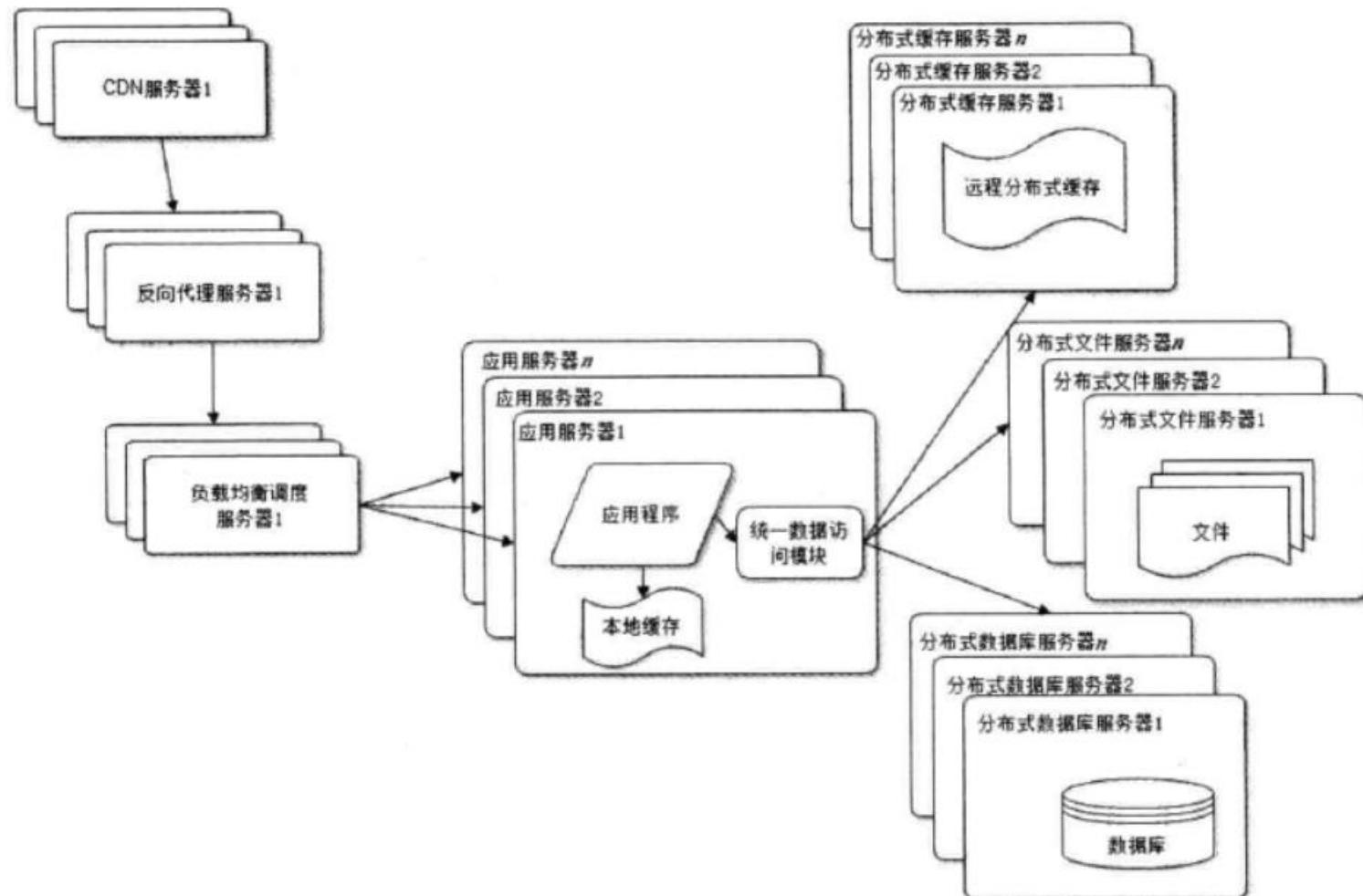
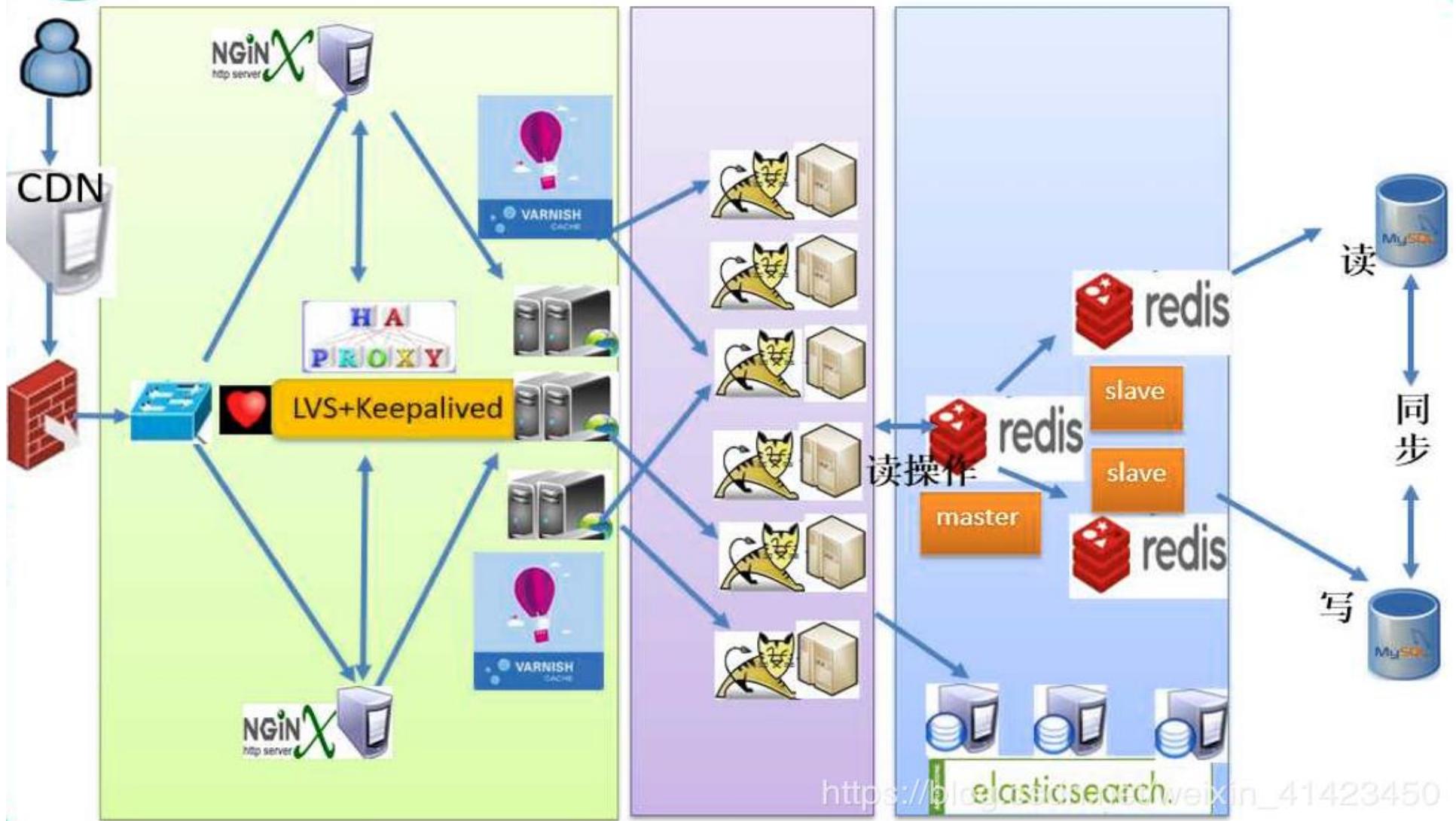


图 1.7 使用分布式文件和分布式数据库系统



1.2.8 使用 NoSQL 和搜索引擎

随着网站业务越来越复杂，对数据存储和检索的需求也越来越复杂，网站需要采用一些非关系数据库技术如 NoSQL 和非数据库查询技术如搜索引擎，如图 1.8 所示。

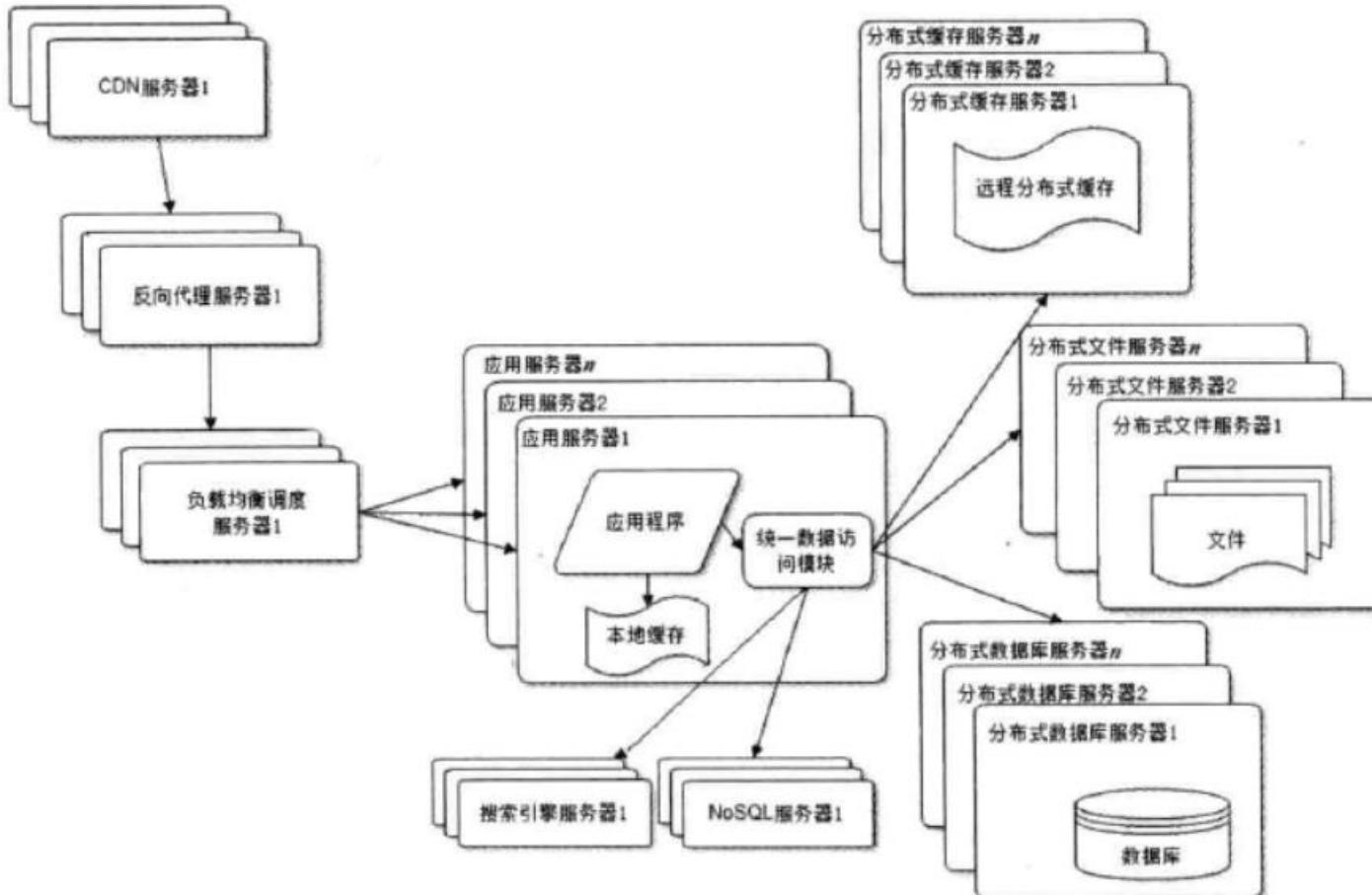


图 1.8 使用 NoSQL 系统和搜索引擎



1.2.9 业务拆分

大型网站为了应对日益复杂的业务场景，通过使用分而治之的手段将整个网站业务分成不同的产品线，如大型购物交易网站就会将首页、商铺、订单、买家、卖家等拆分成不同的产品线，分归不同的业务团队负责。

具体到技术上，也会根据产品线划分，将一个网站拆分成许多不同的应用，每个应用独立部署维护。应用之间可以通过一个超链接建立关系（在首页上的导航链接每个都指向不同的应用地址），也可以通过消息队列进行数据分发，当然最多的还是通过访问同一个数据存储系统来构成一个关联的完整系统，如图 1.9 所示。

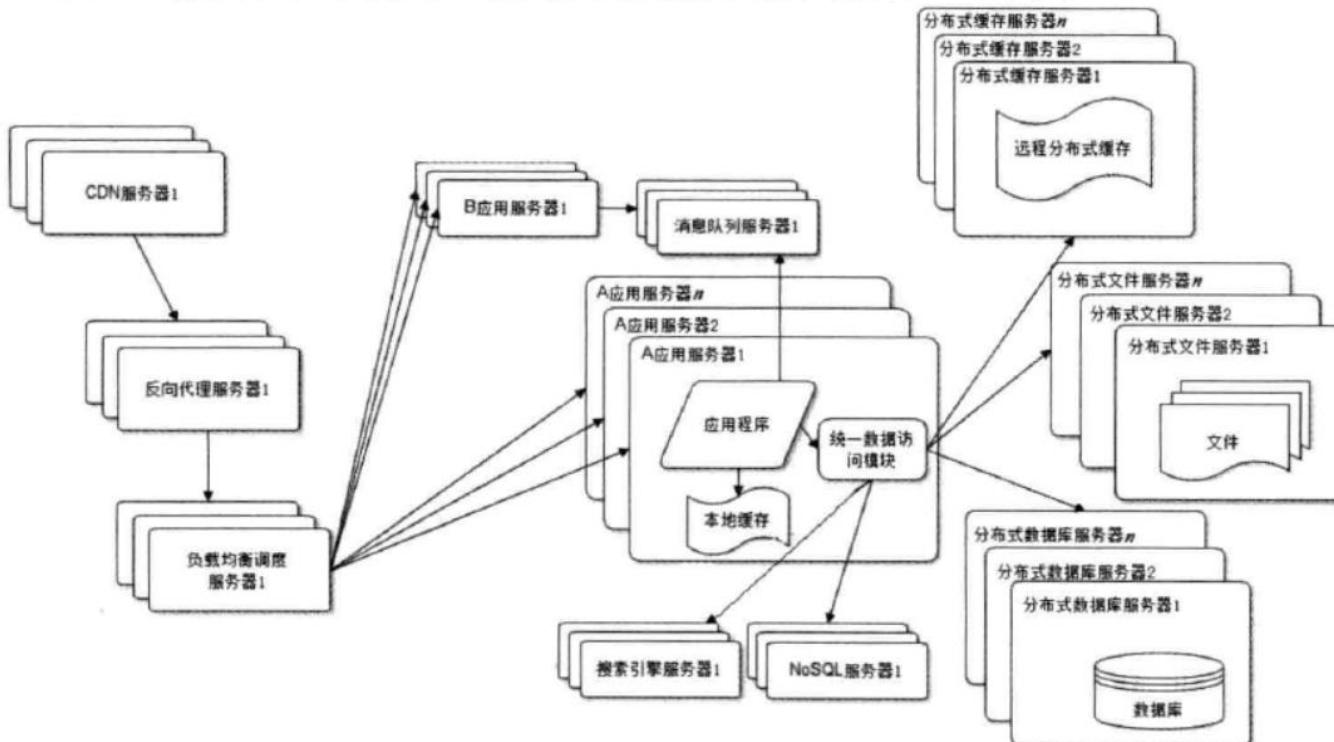


图 1.9 应用拆分

1.2.10 分布式服务

随着业务拆分越来越小，存储系统越来越庞大，应用系统的整体复杂度呈指数级增加，部署维护越来越困难。由于所有应用要和所有数据库系统连接，在数万台服务器规模的网站中，这些连接的数目是服务器规模的平方，导致存数据库接资源不足，拒绝服务。

既然每一个应用系统都需要执行许多相同的业务操作，比如用户管理、商品管理等，那么可以将这些共用的业务提取出来，独立部署。由这些可复用的业务连接数据库，提供共用业务服务，而应用系统只需要管理用户界面，通过分布式服务调用共用业务服务完成具体业务操作，如图 1.10 所示。

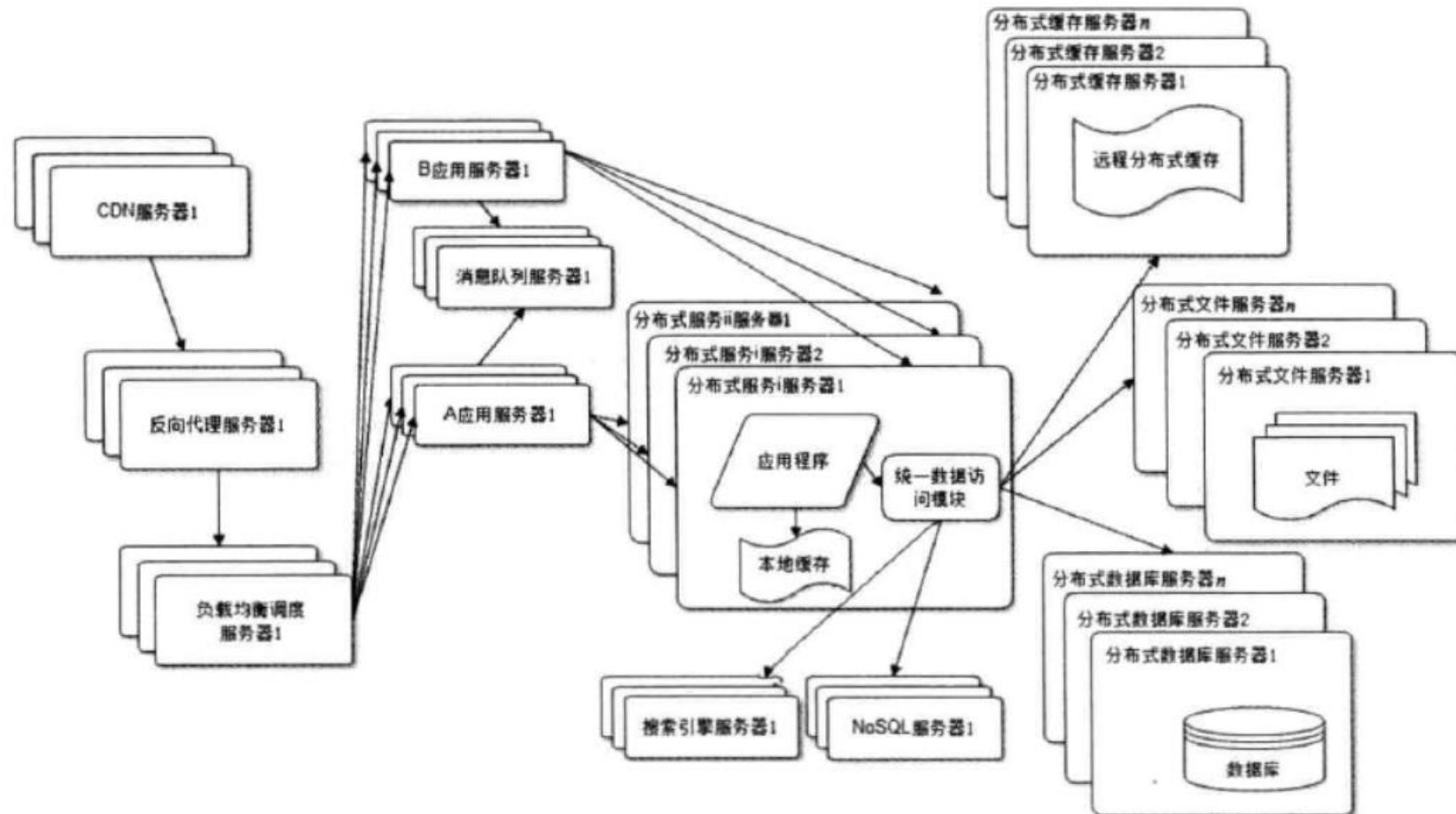
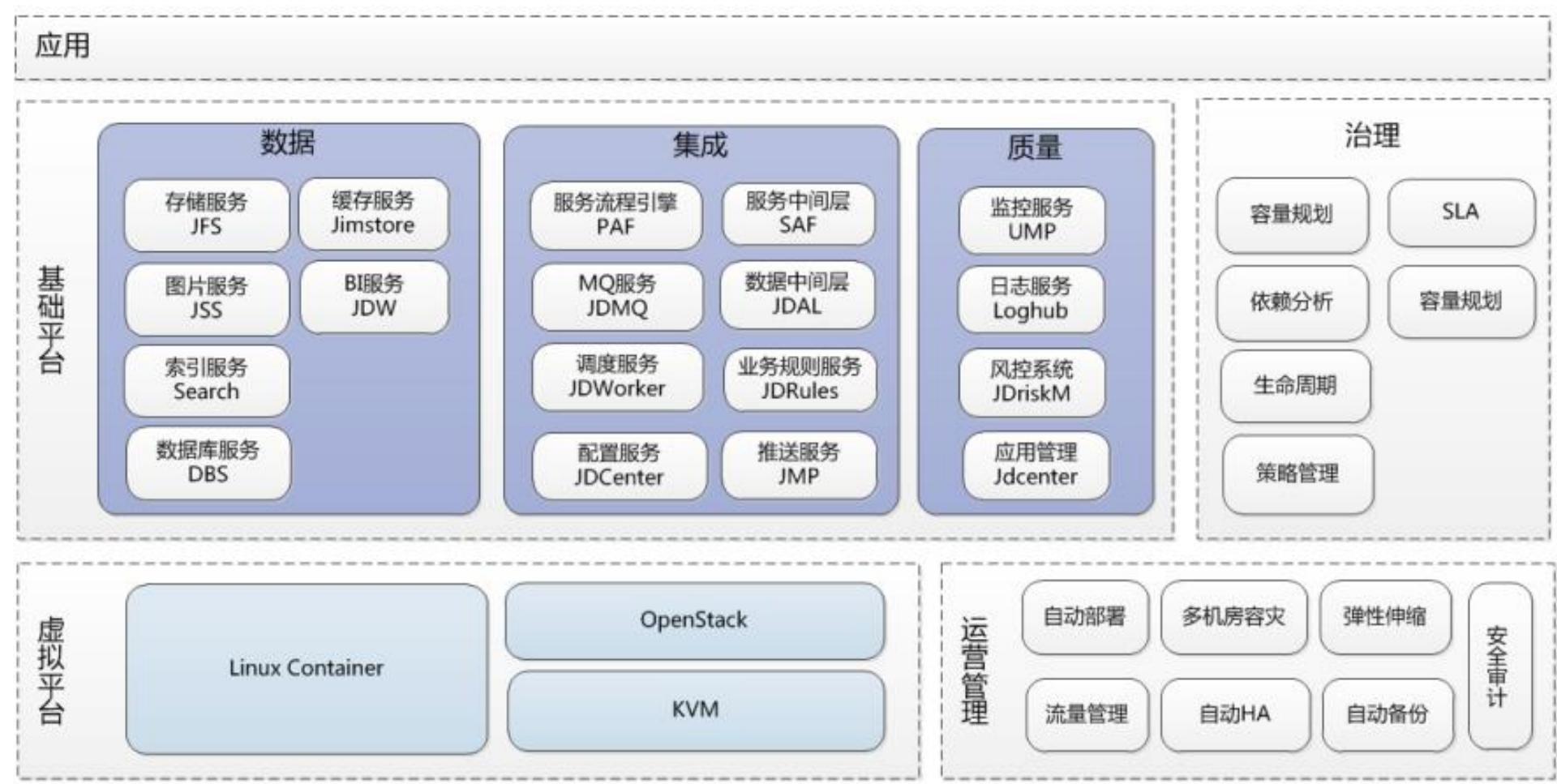
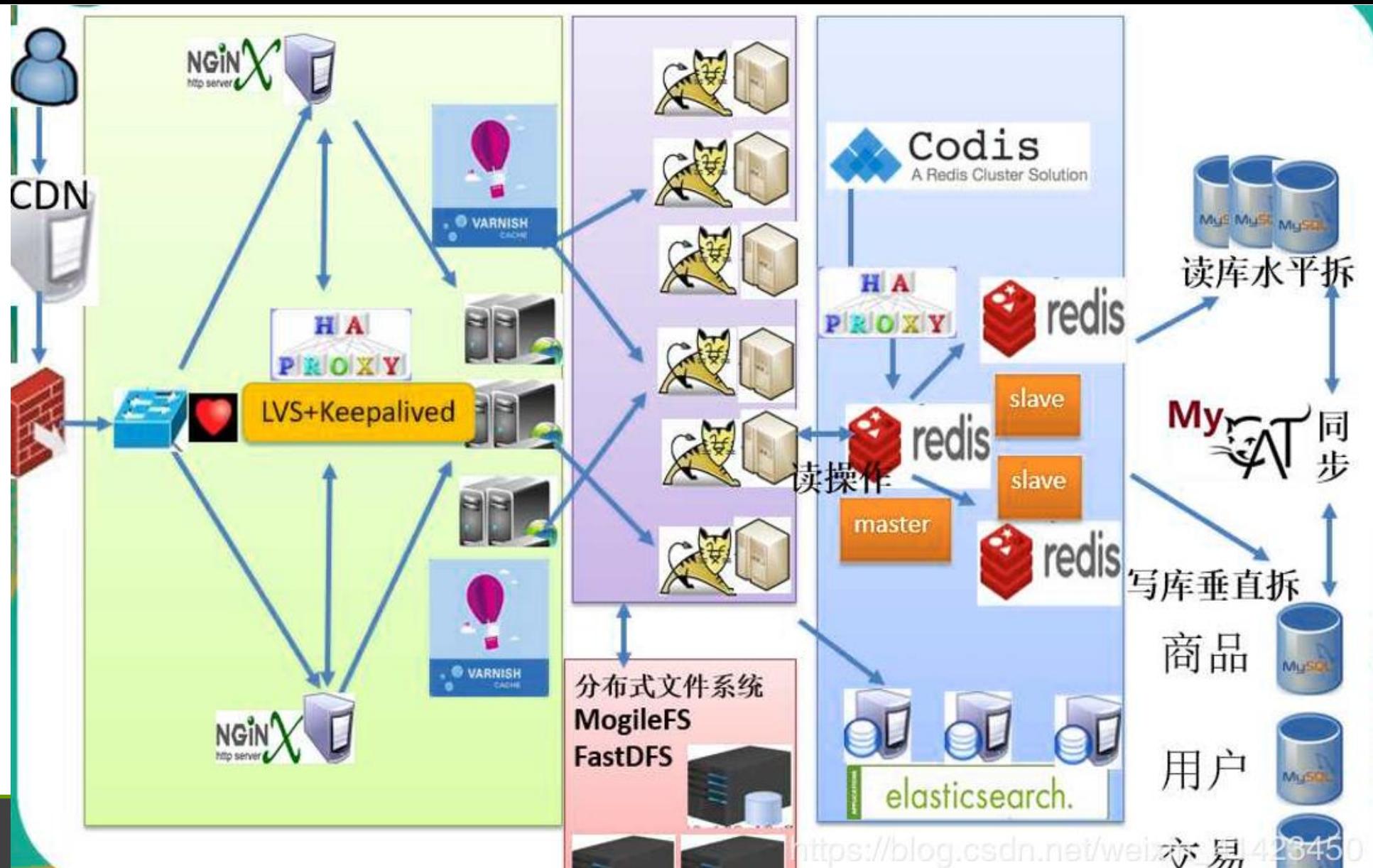


图 1.10 分布式服务

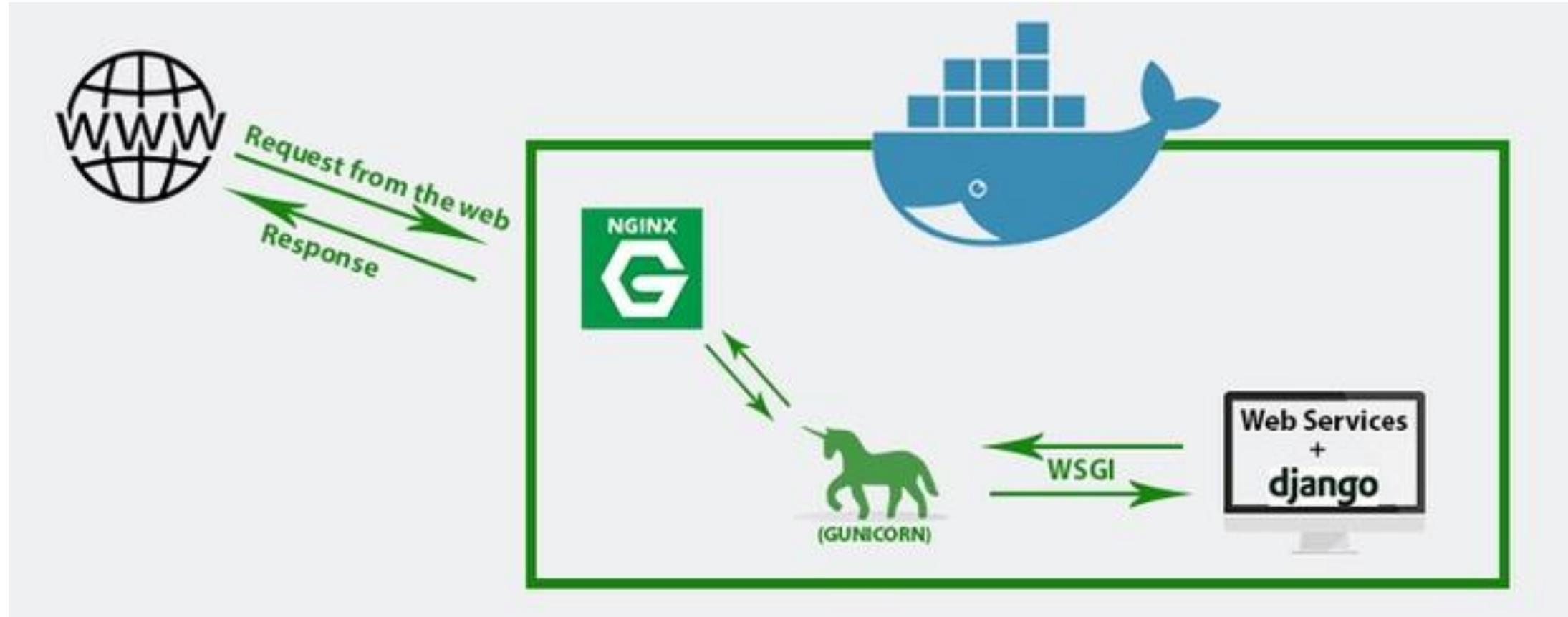
大型网站的架构演化到这里，基本上大多数的技术问题都得以解决，诸如跨数据中心的实时数据同步和具体网站业务相关的问题也都可以通过组合改进现有技术架构来解决。

京东基础技术架构：

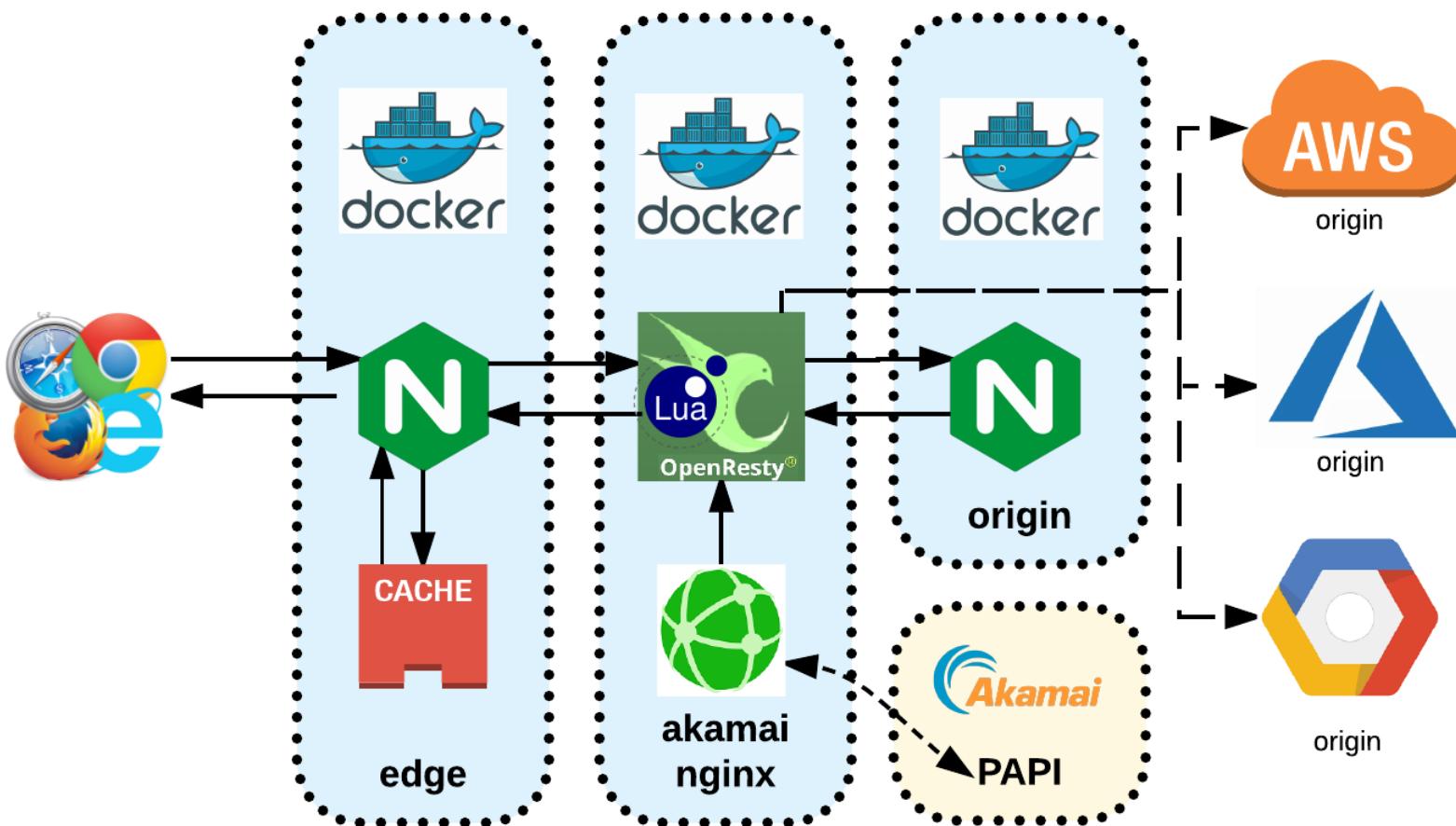




现在很流行Cloud 上部署 – 但， 架构还是类似



akamai-nginx request flow



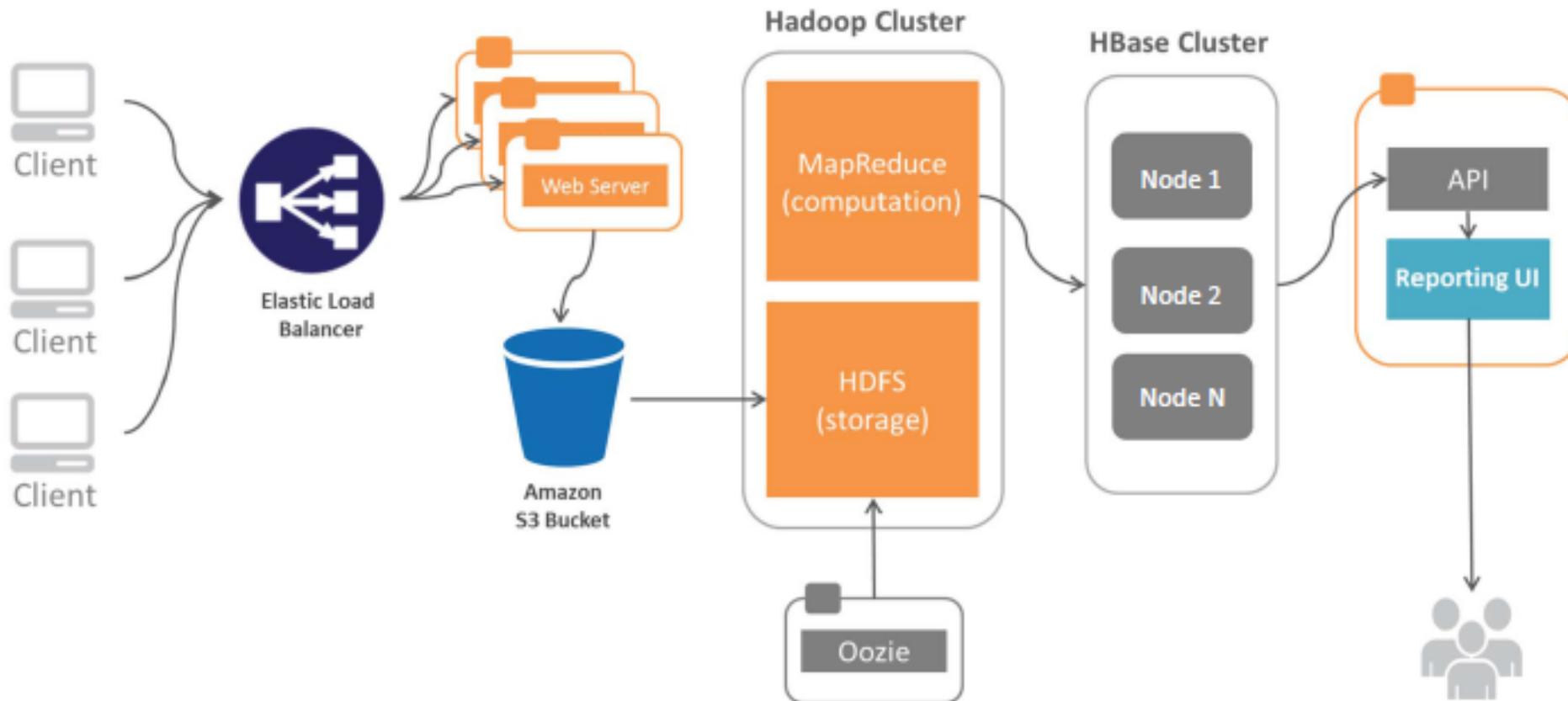
Akamai PAPI is called once, to generate equivalent LUA config

□ Many Cloud platforms



Technologies:

- Amazon S3
- Flume
- Hadoop/HDFS, MapReduce
- HBase
- Oozie
- Hive



Summary

业务系统(Business System) 是立身之本~

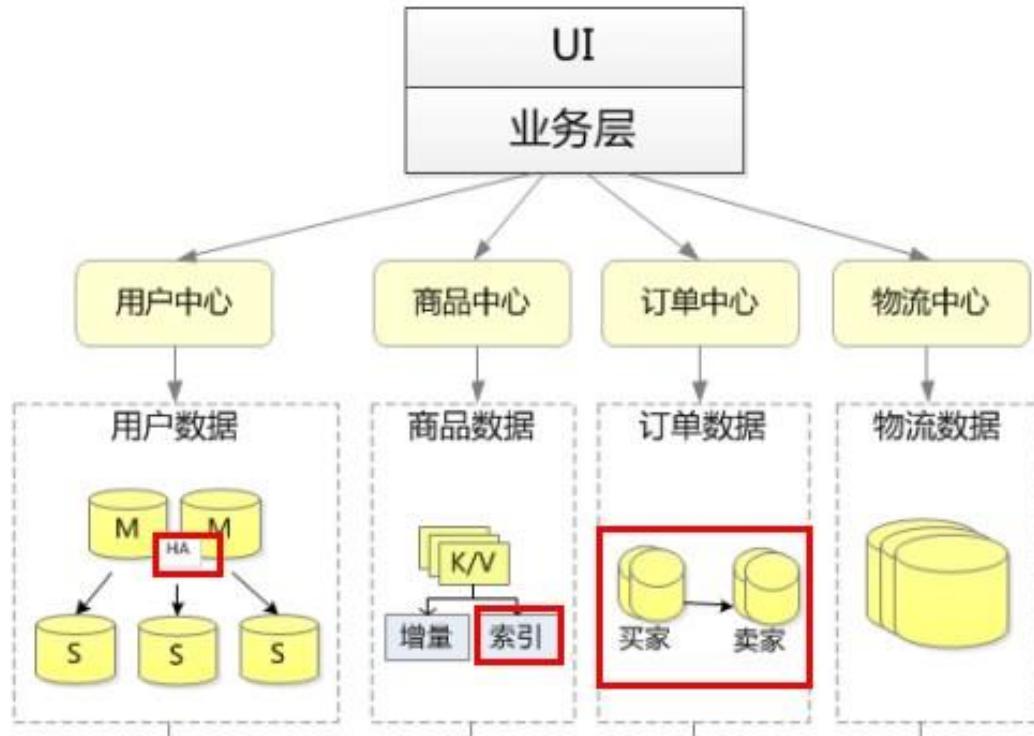
业务系统
(如，如何支持订单？
极限情况下的订单处理？)

数据分析

数据的采集和保存
(如，围绕订单的处理
而需要维护的数据)

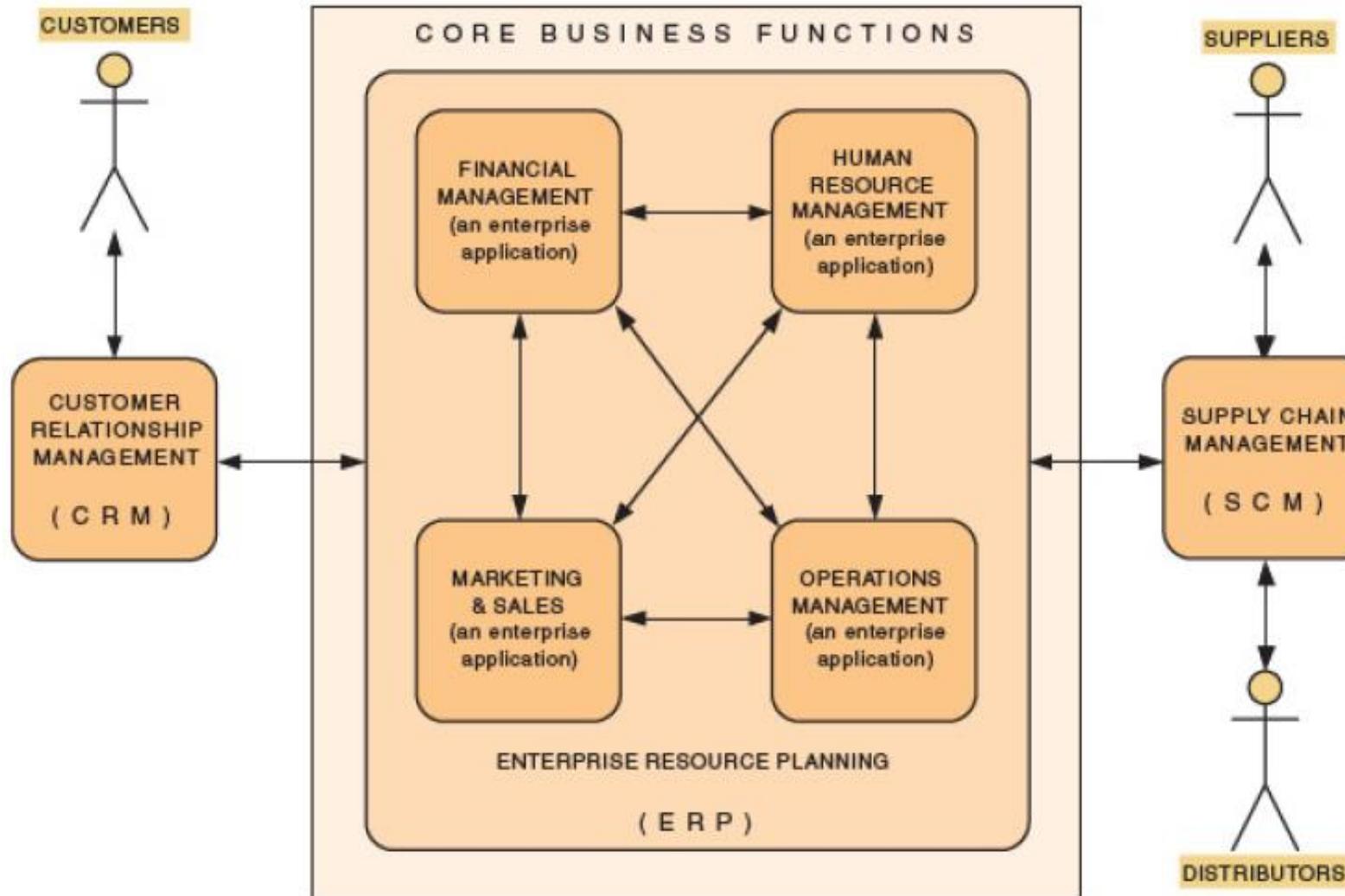


业务系统(Business System) 是立身之本~

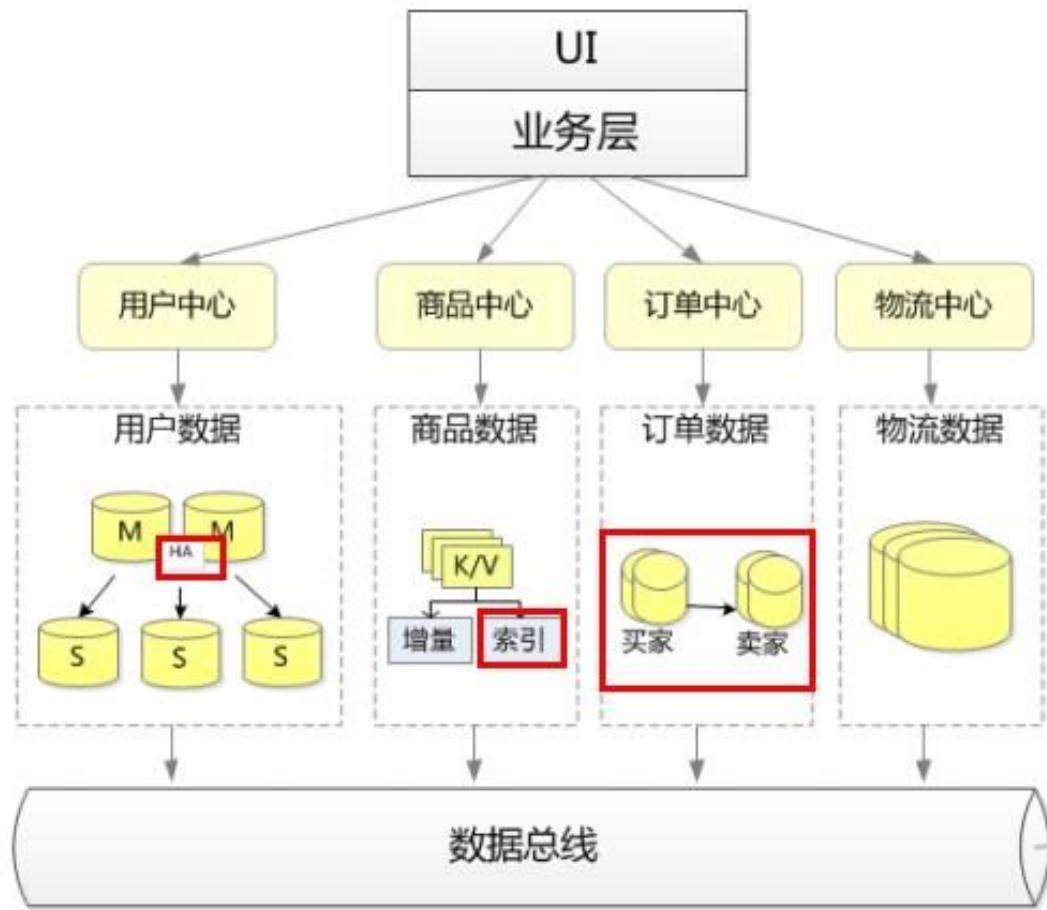


数据分析

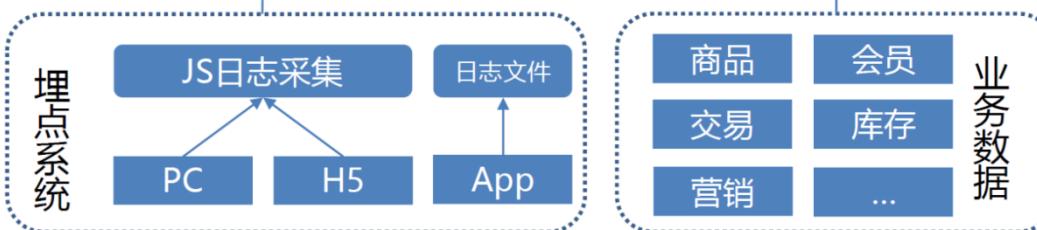
数据的采集和保存
(如，围绕订单的处理
而需要维护的数据)



业务系统(Business System) 是立身之本~

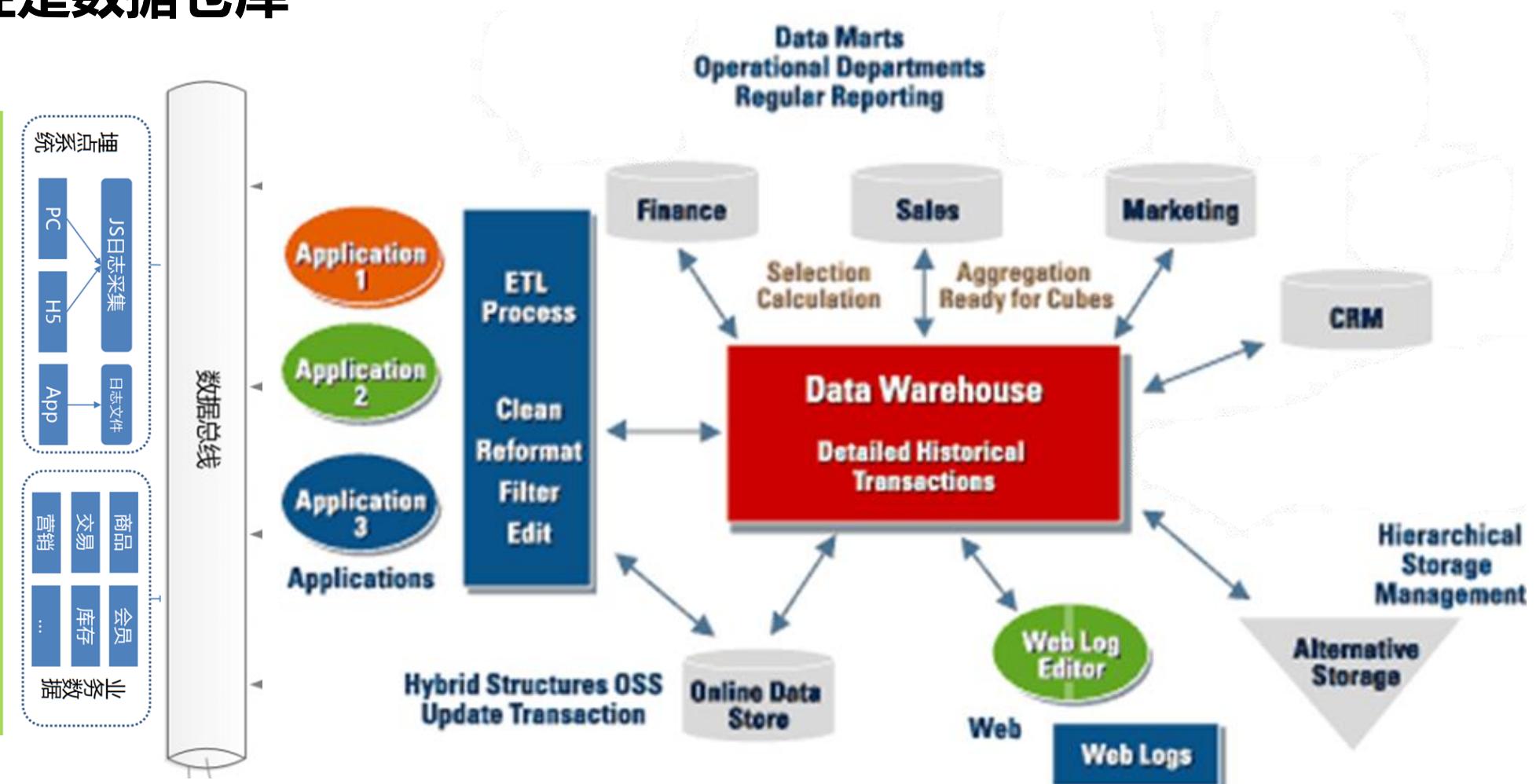


数据分析

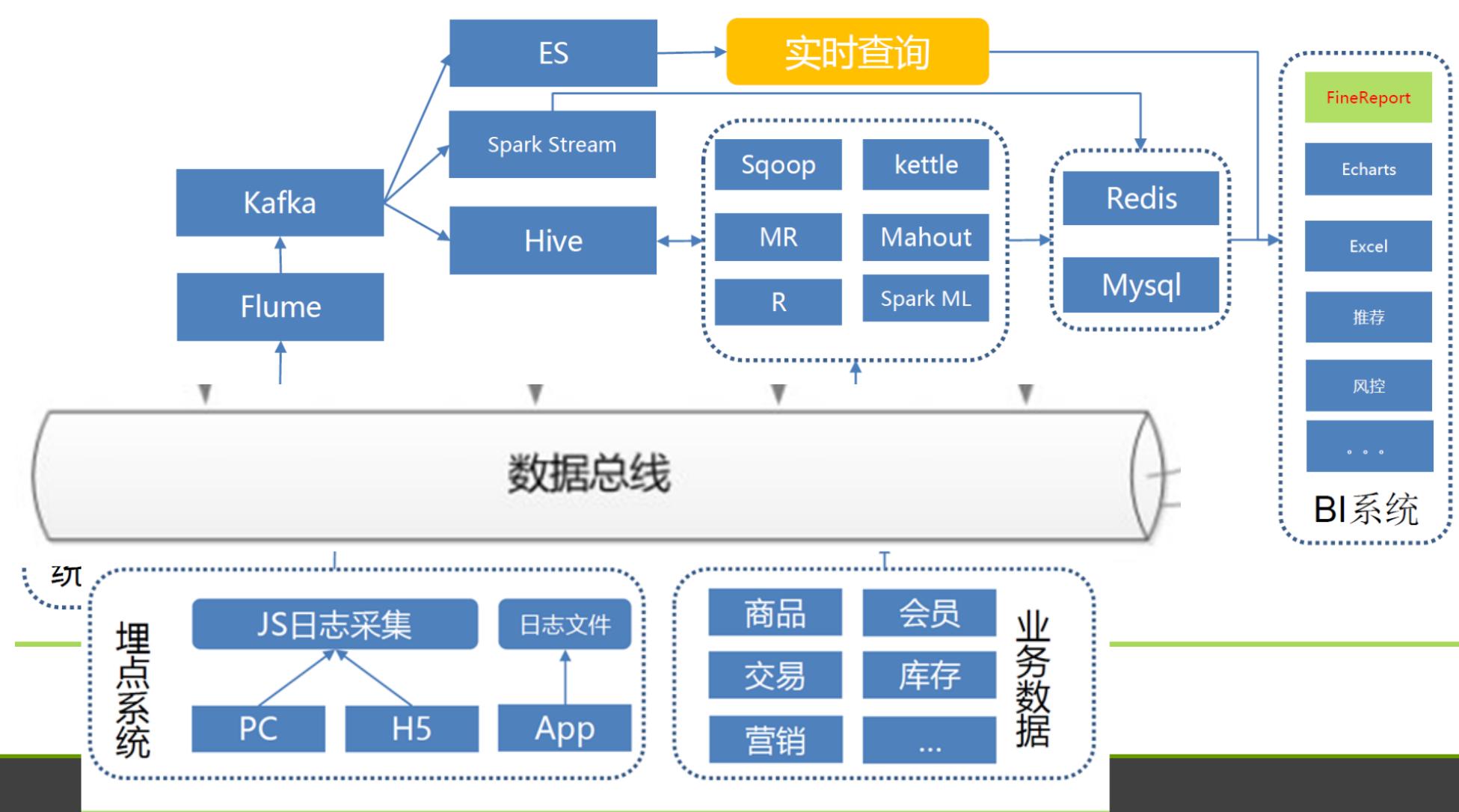


(大)数据分析([Big]Data Analytics) 算是增值吧~

□ 曾经是数据仓库



□ 现在是大数据分析

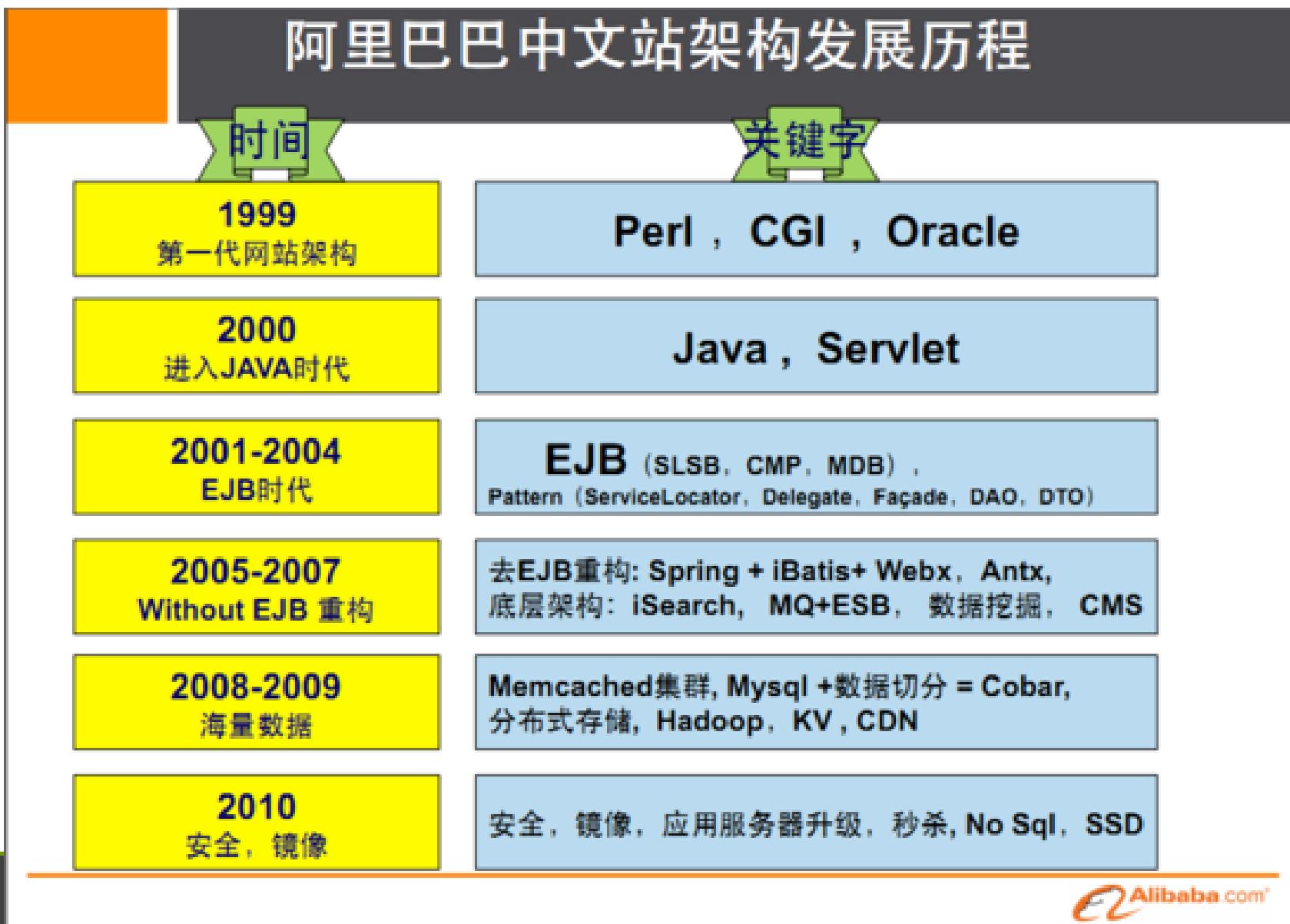


某(大)数据分析创业公司的架构



架构分解就是为了满足高并发和大数据，具体原则：



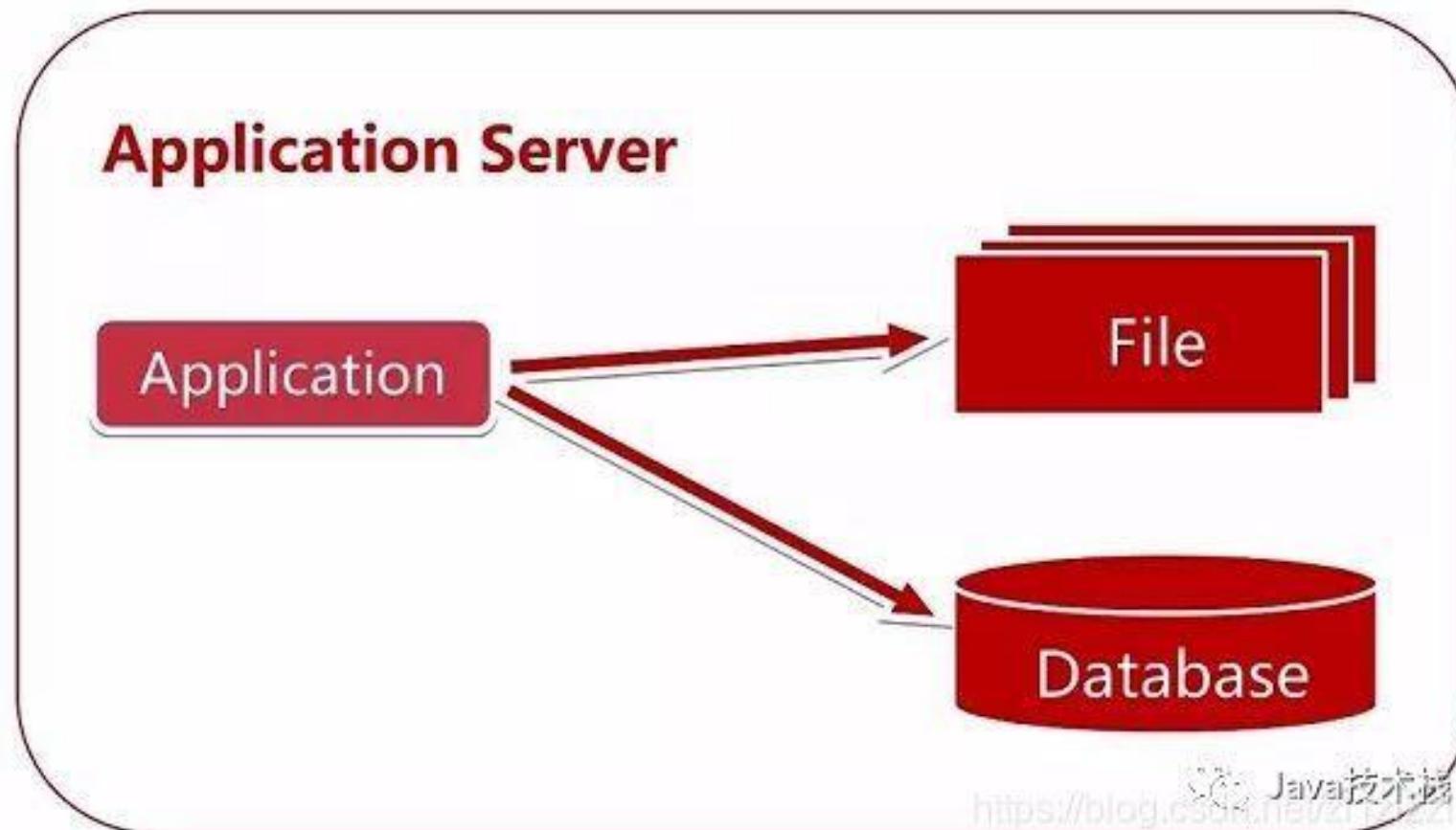


以淘宝为例，解析大型电商服务端架构！

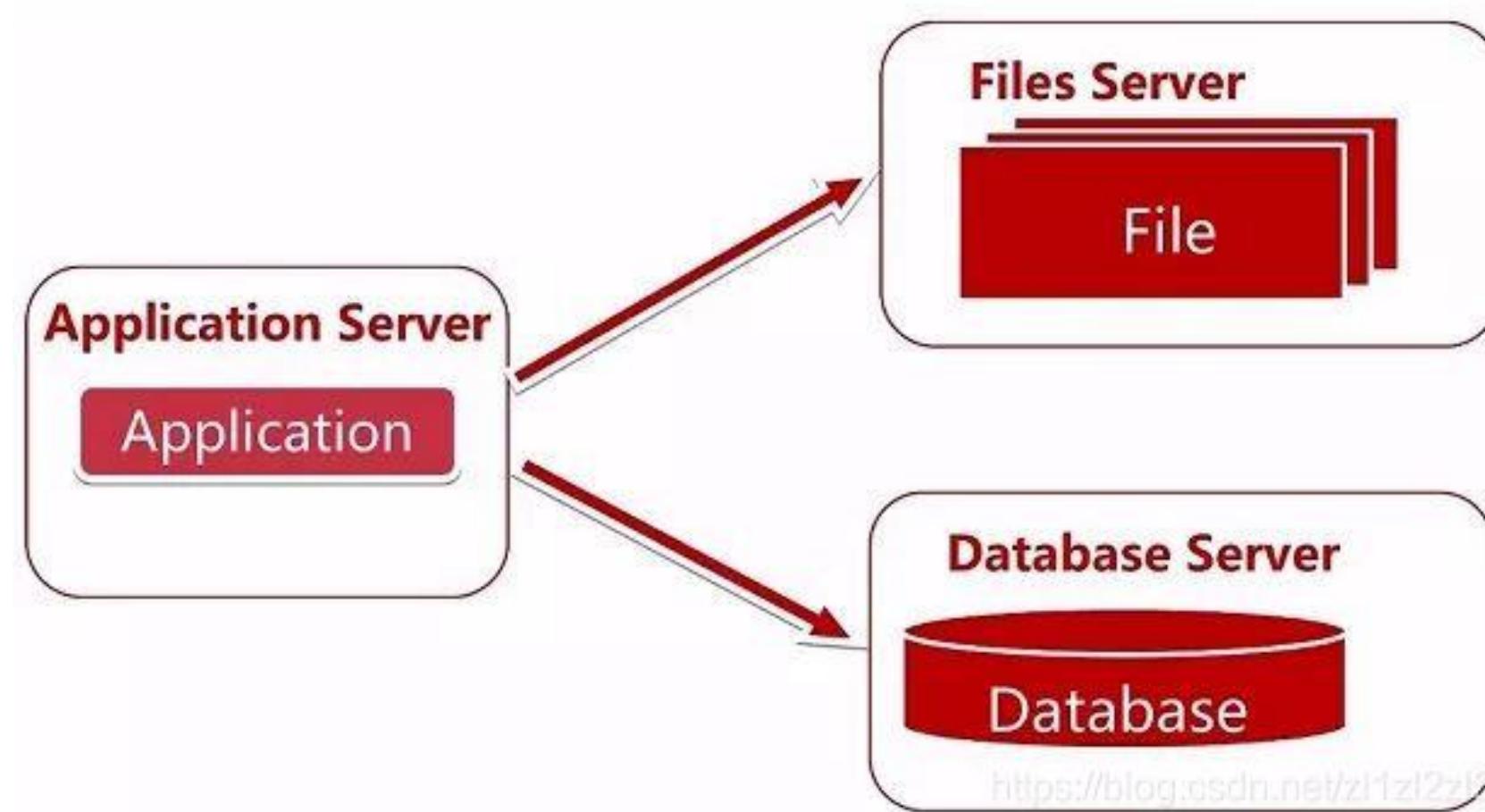
<https://blog.csdn.net/zl1zl2zl3/article/details/105126660>



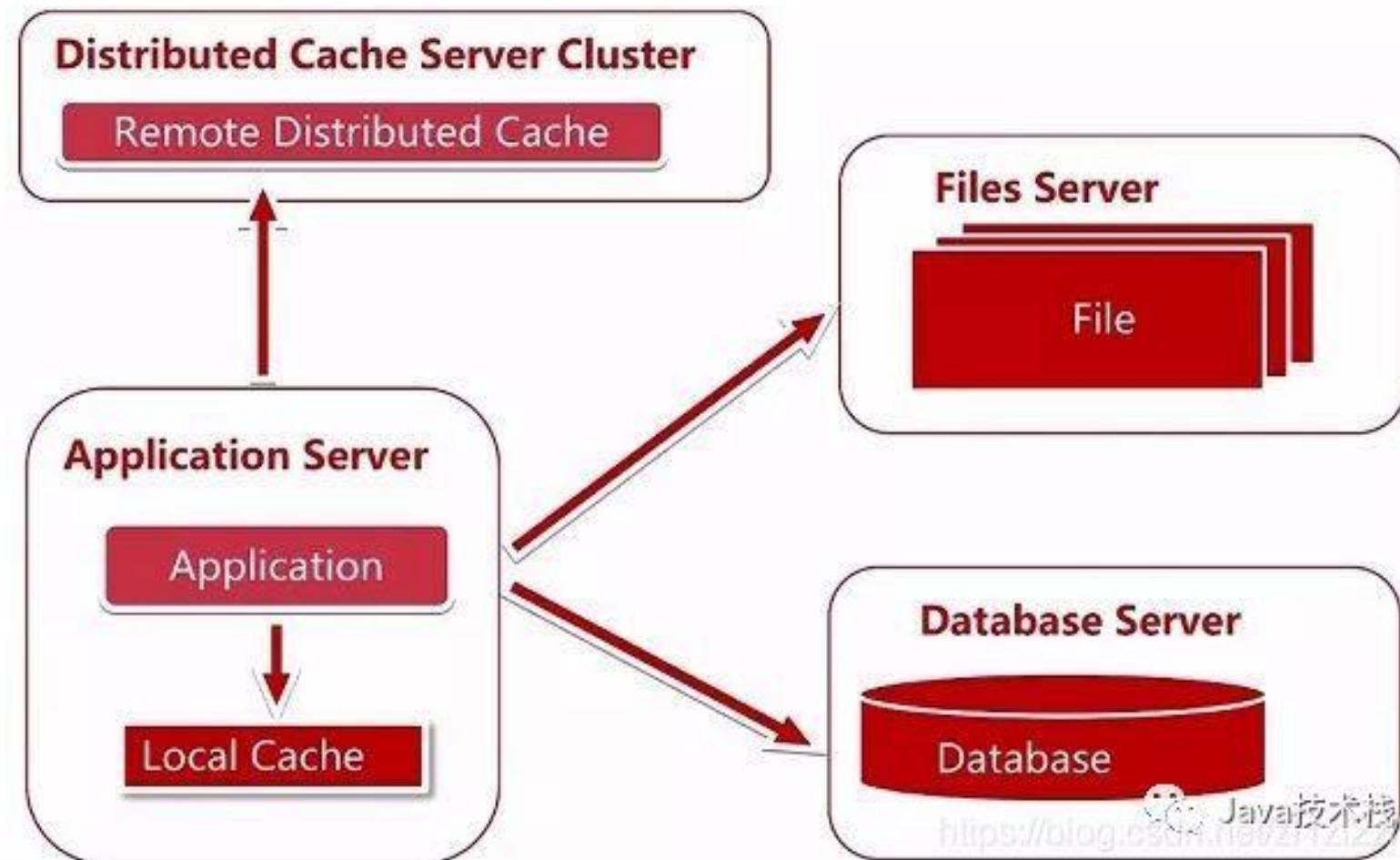
□ 单机架构



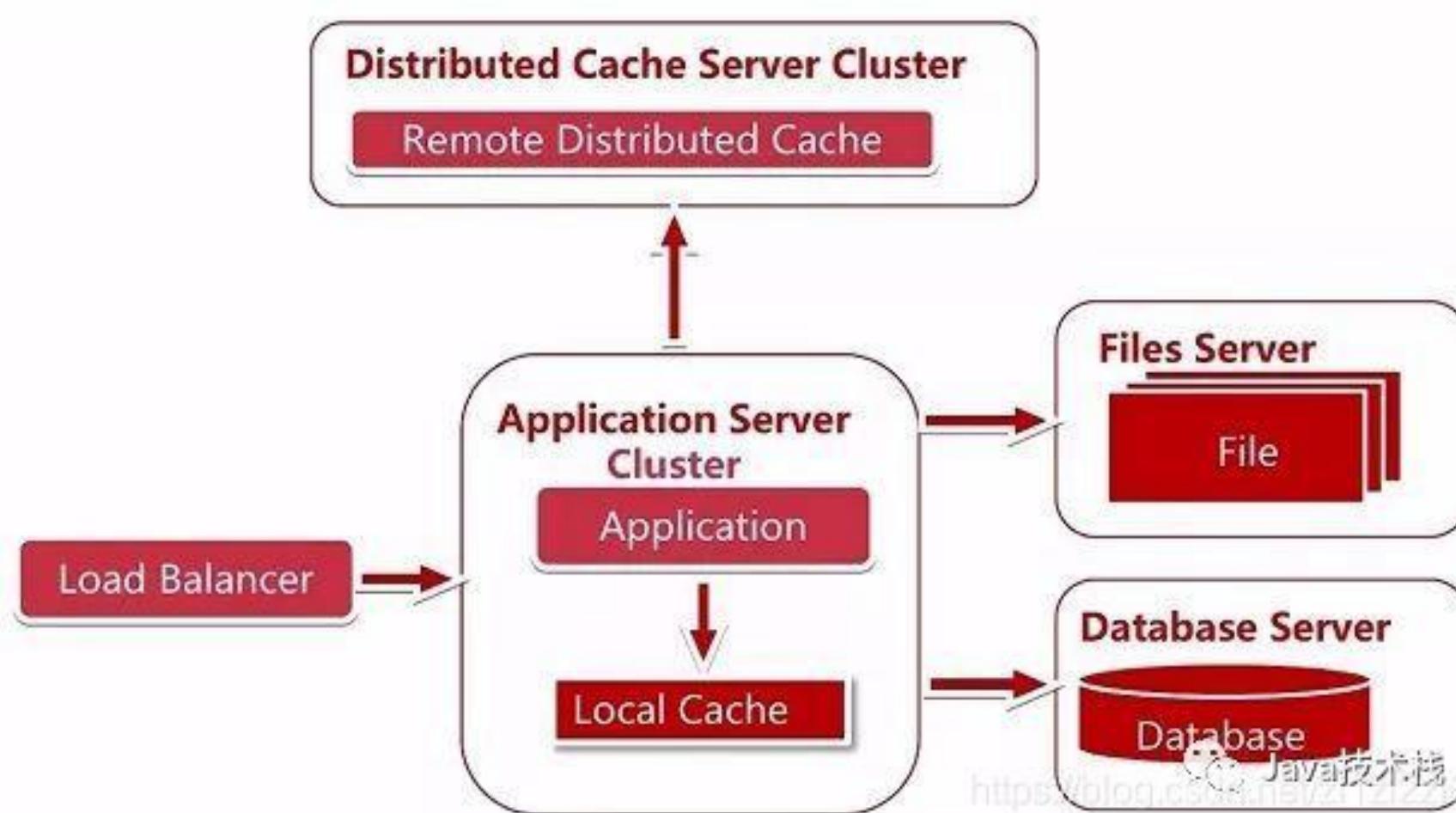
□ 多机部署



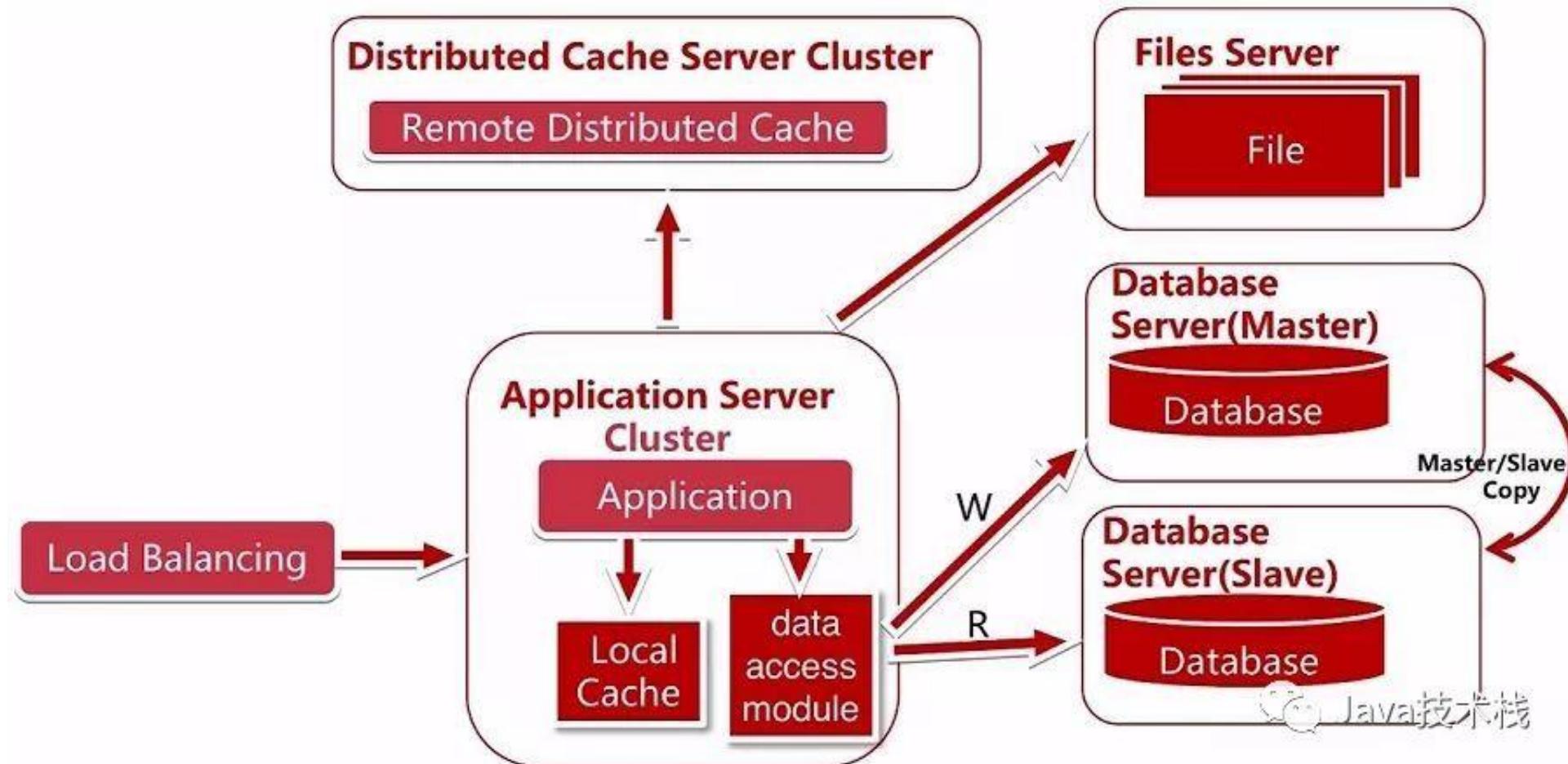
□ 分布式缓存



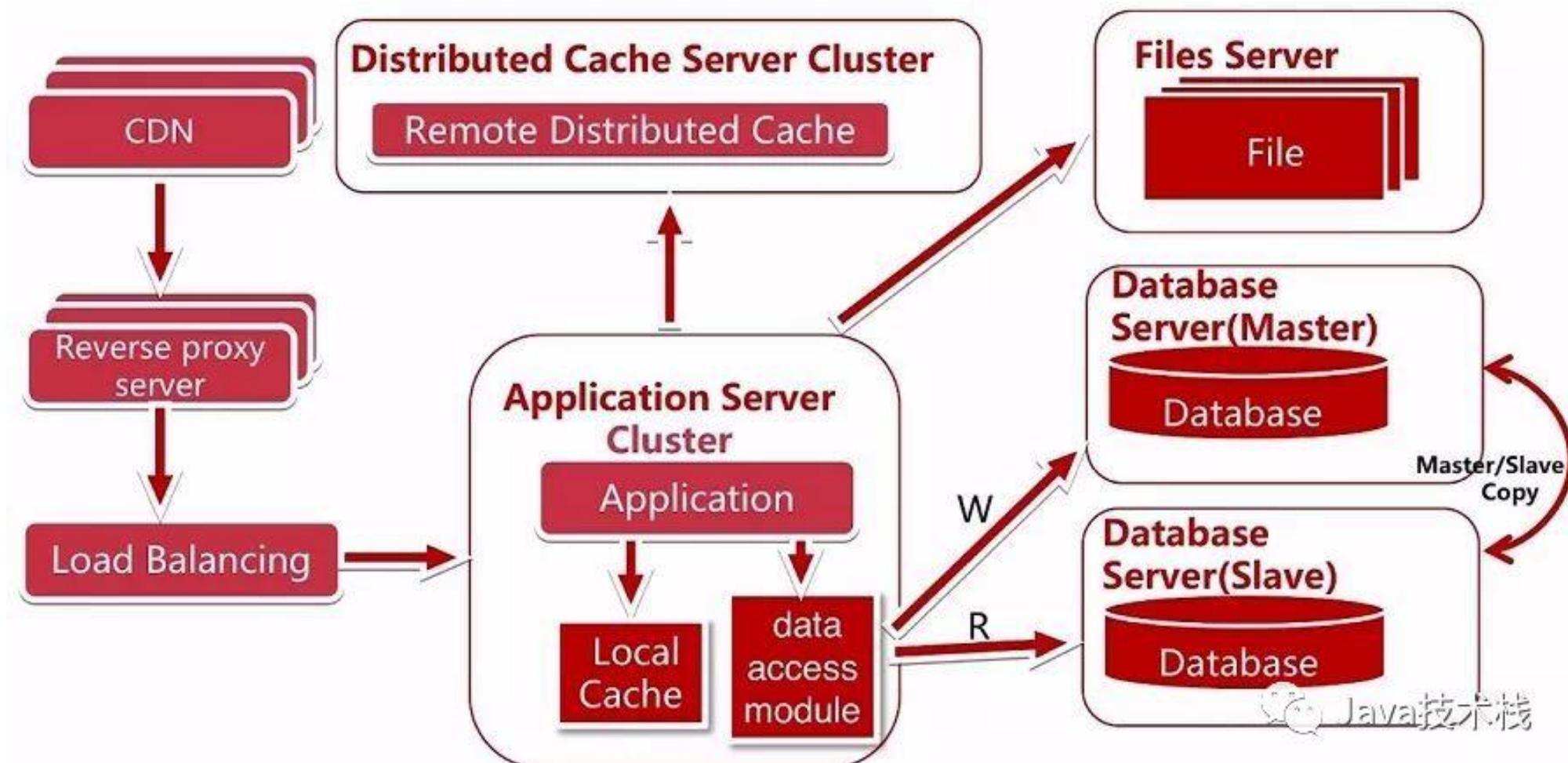
□ 服务器集群



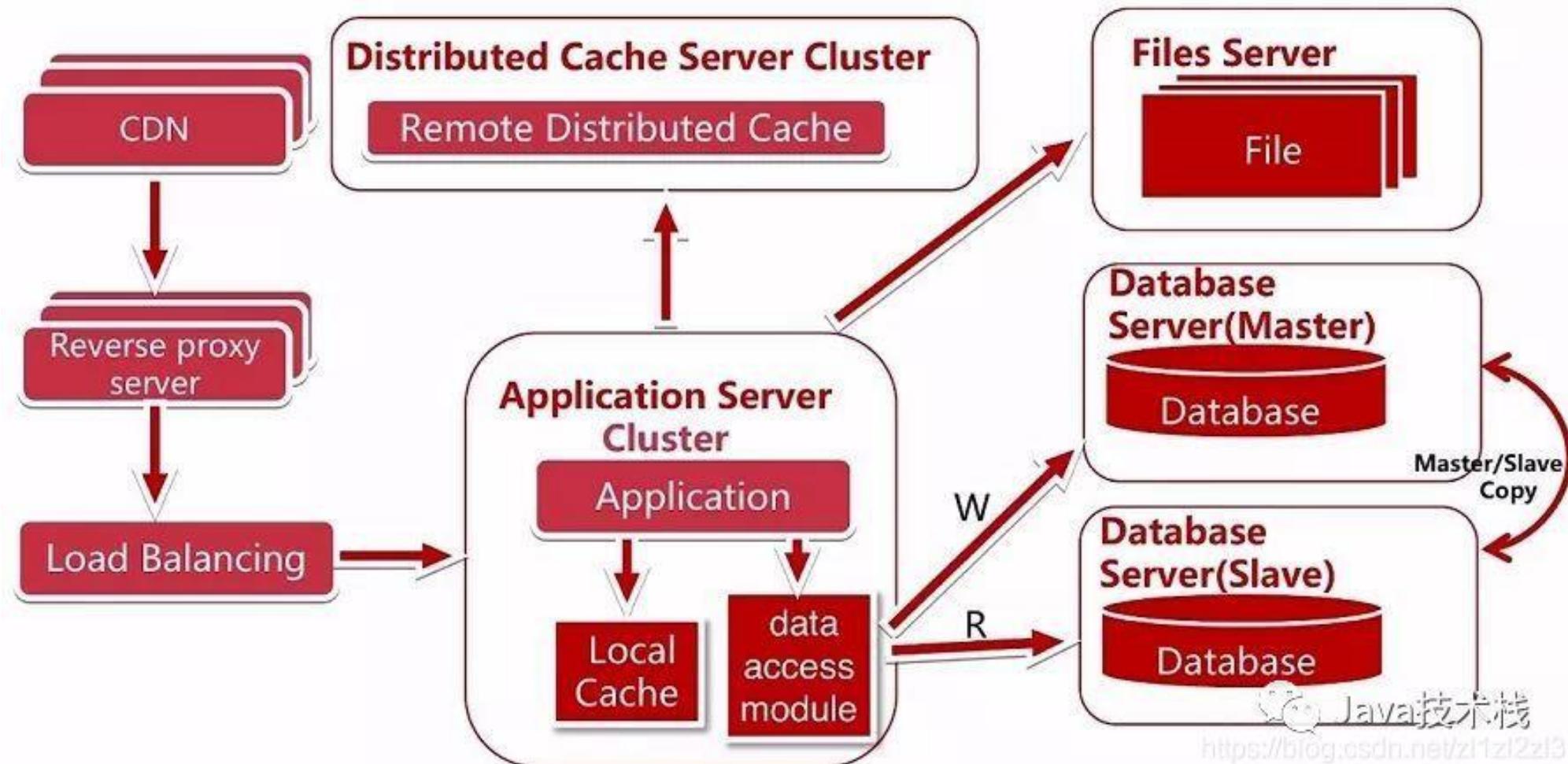
□ 数据库读写分离



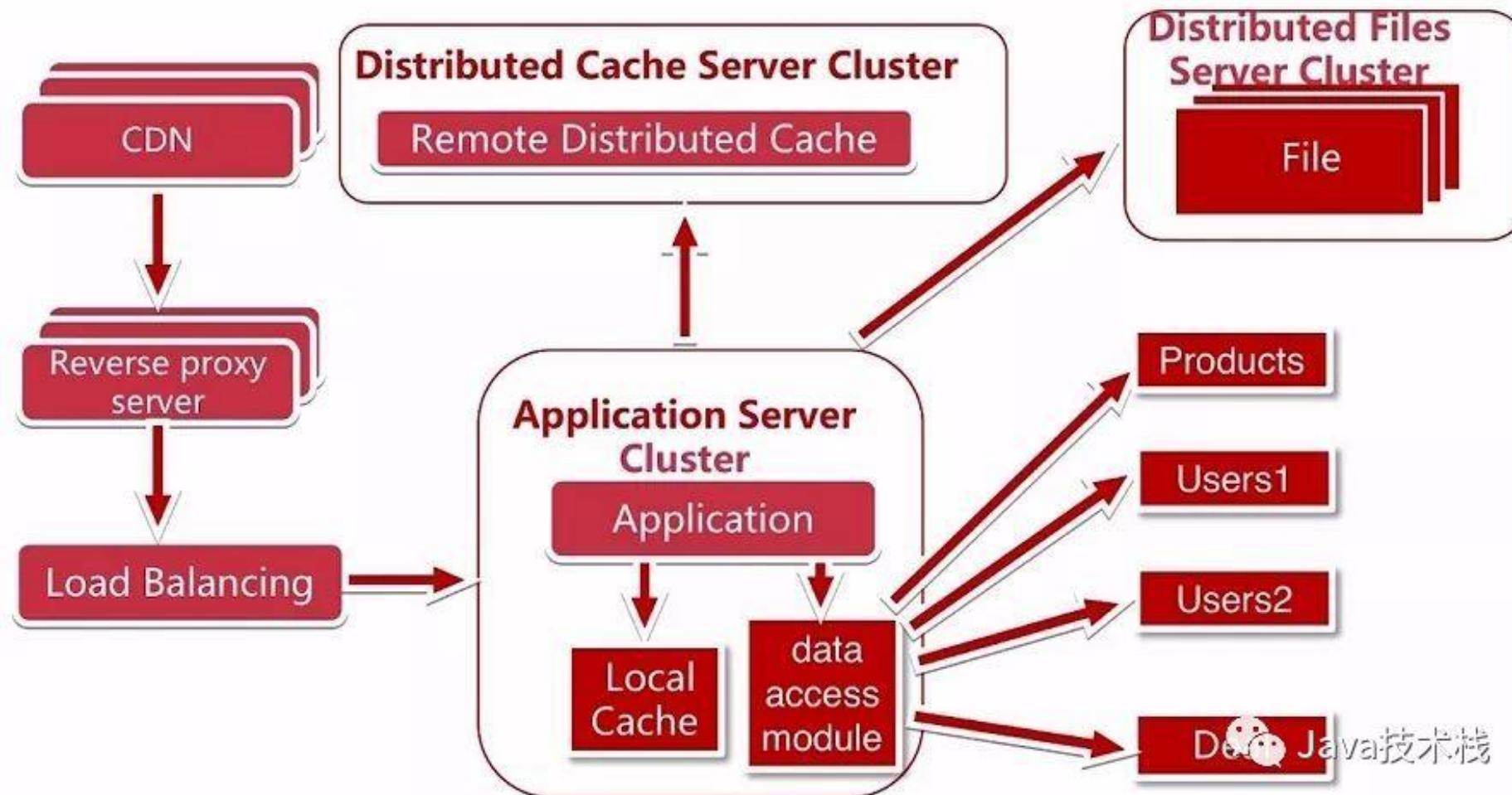
□ CDN 加速与反向代理



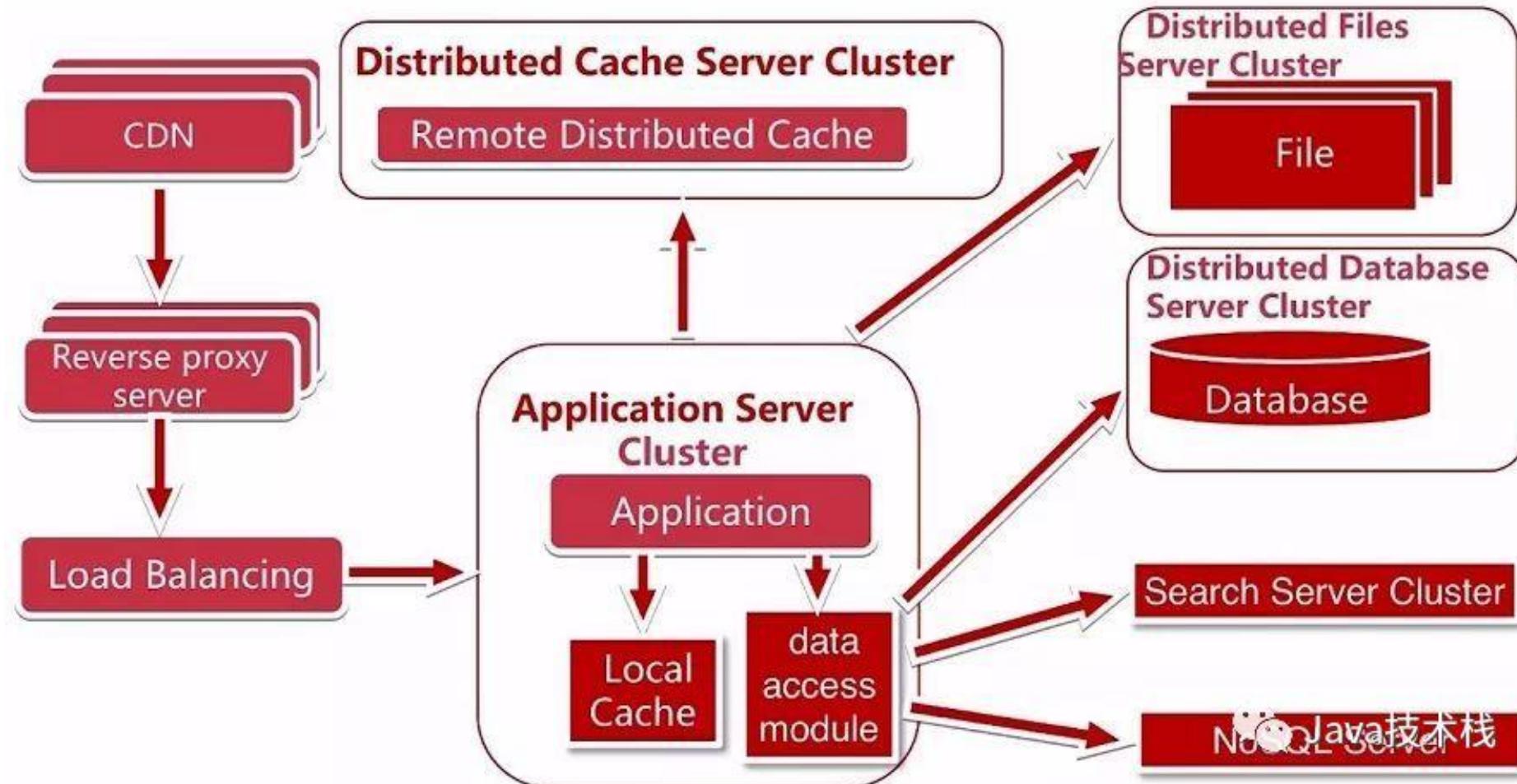
分布式文件服务器



□ 数据库分库分表



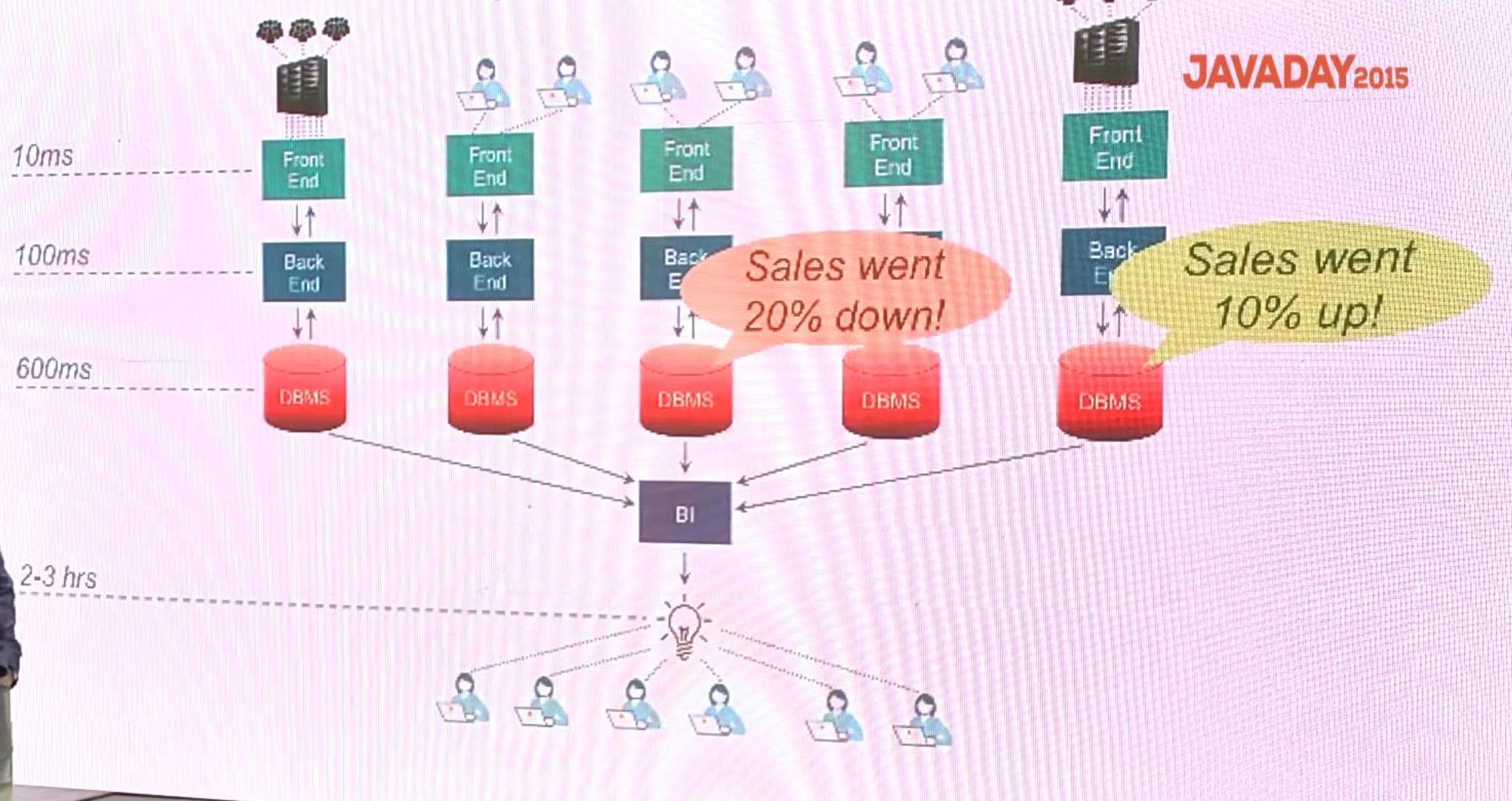
□ 搜索引擎与 NoSQL



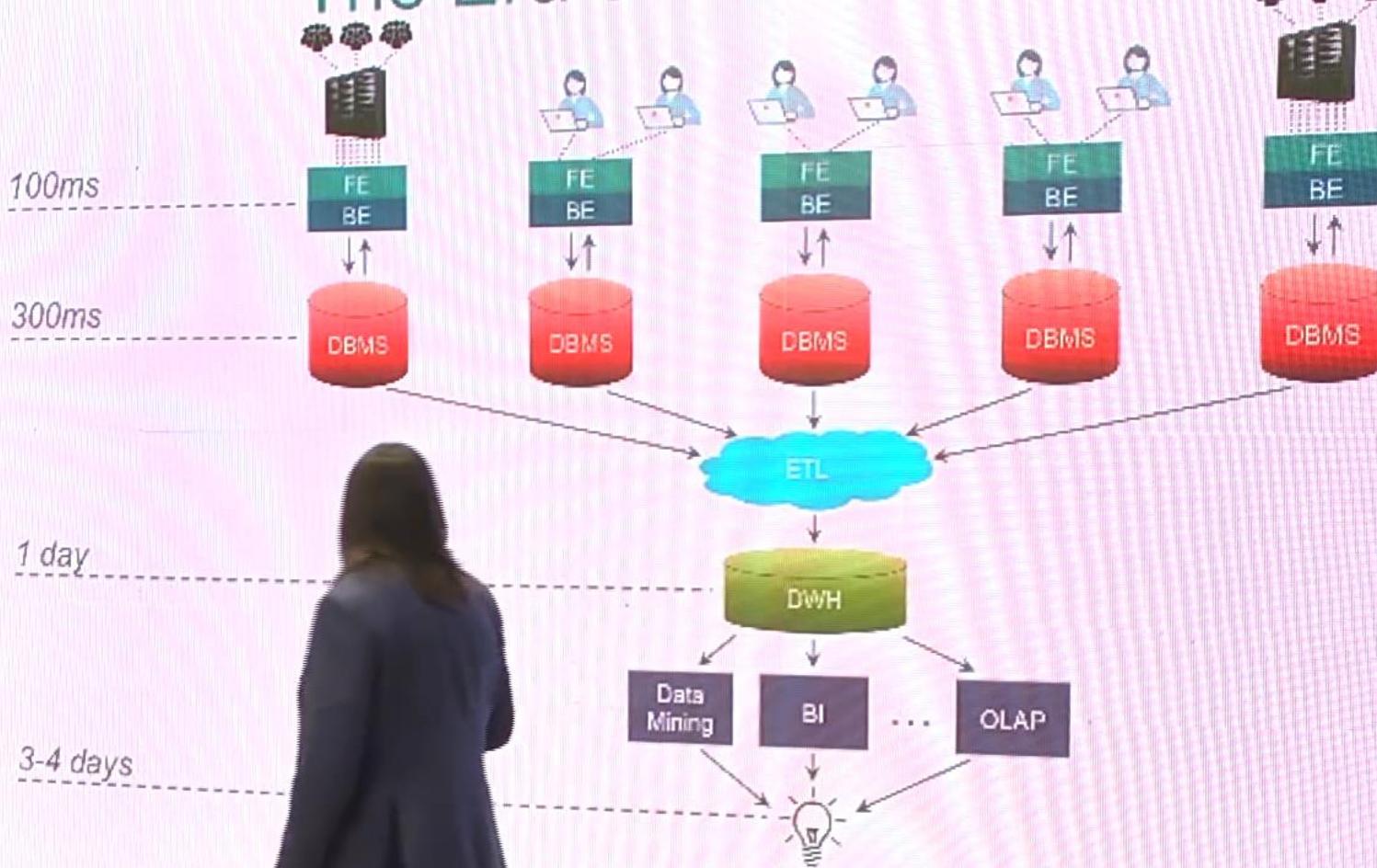
□ 2020年7月26日19:03:10

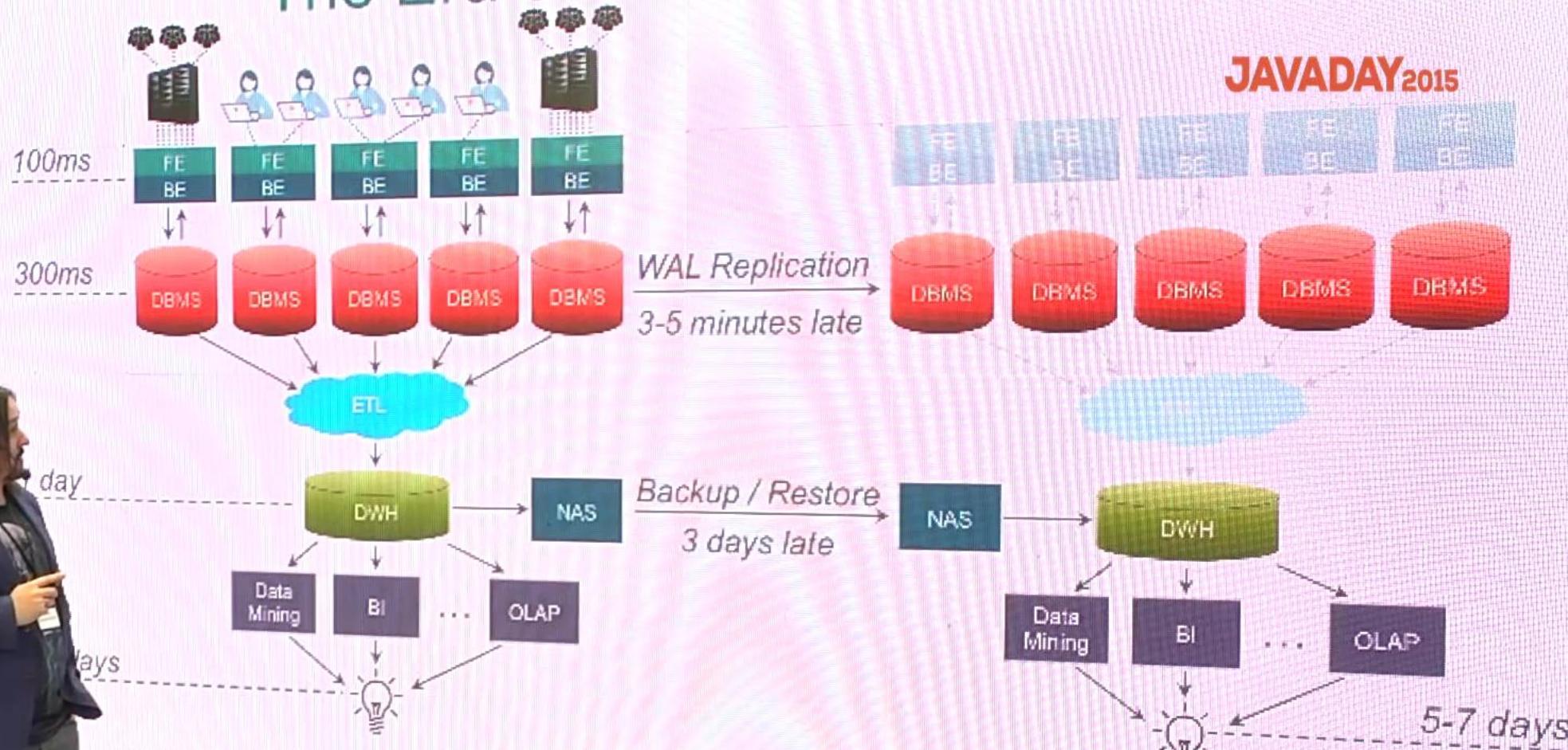
□ 整理一些视频，看到 JavaDay 2015 关于数据处理演变的报告，涉及到如何通过Split 应对大数据、高并发的技术演变，摘记一下





A man with long hair and a beard, wearing a dark blue jacket and grey pants, stands on a stage, gesturing towards the presentation screen.





Java Day 2015
JUG.UA

Java
Day 2015

LUXOFT



ORACLE

DATAART

Cogniance

GlobalLogic

SIGMA
Software

accenture

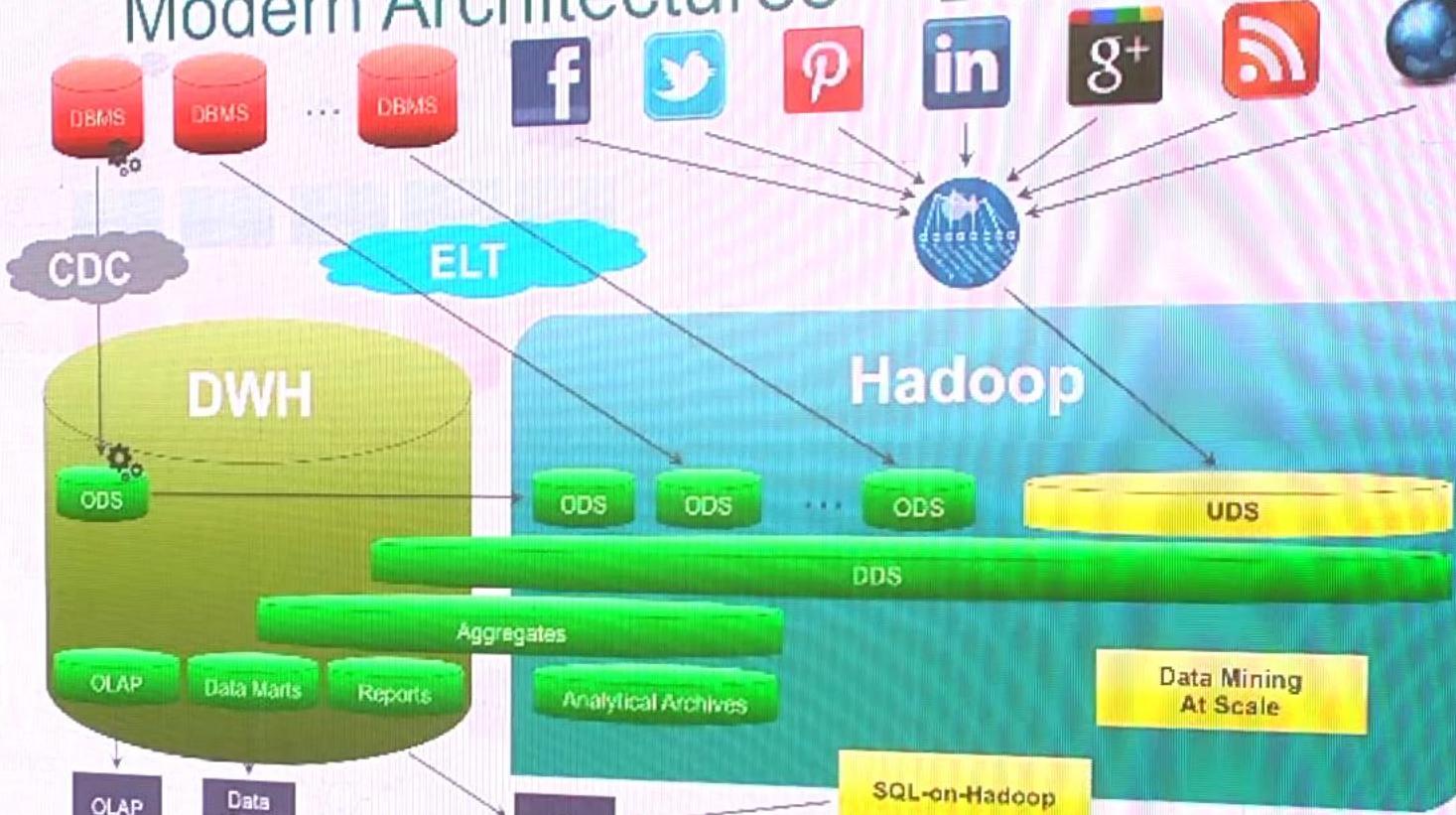
BRAINS

Jelastic

Playtika

INTRIG PRO

Modern Architectures – Data Lake

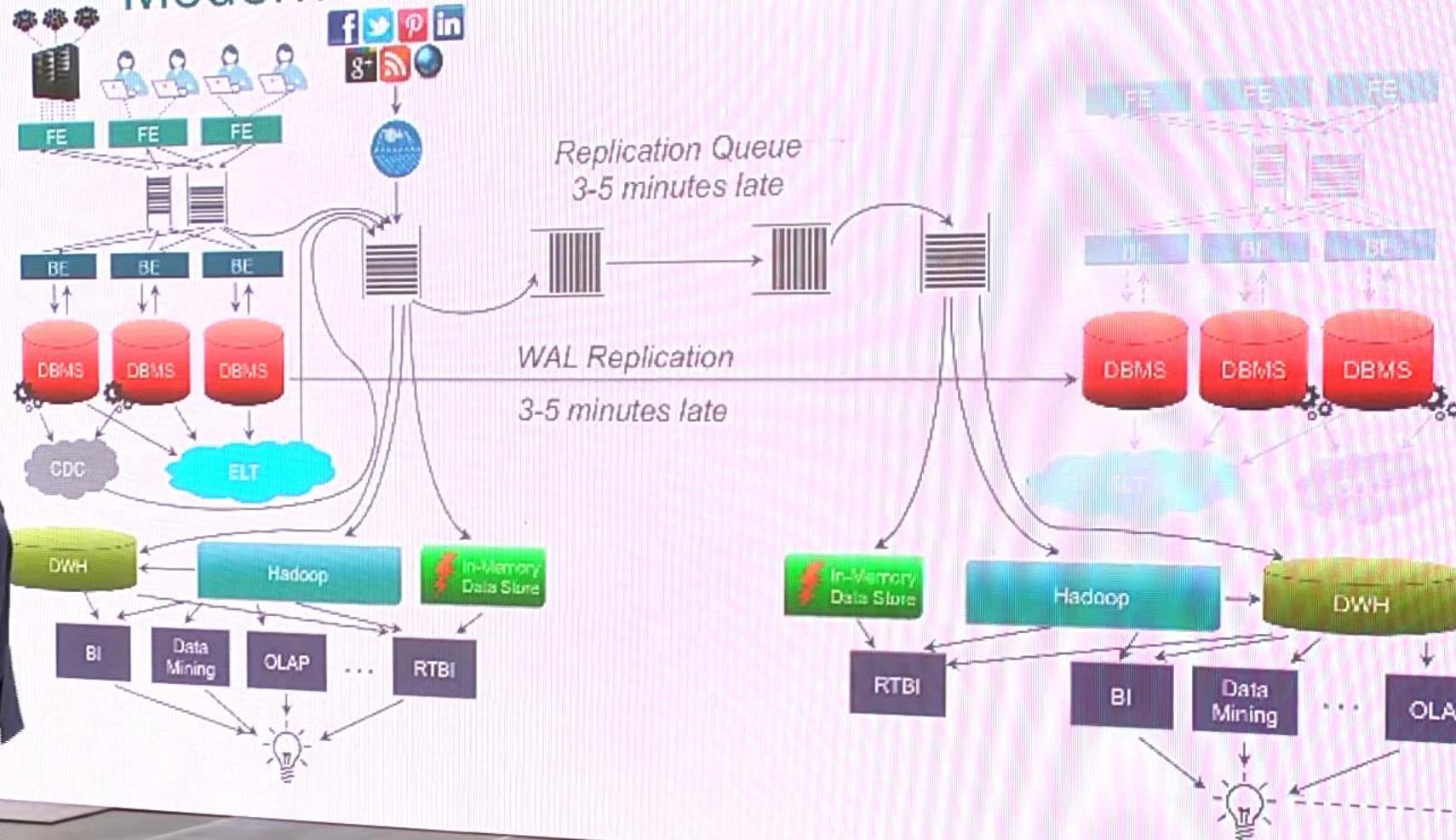


A man in a dark suit and glasses stands on stage, speaking into a microphone. He is positioned to the left of a large projection screen displaying the architectural diagram. To his left is a white podium with a small Java Day 2015 logo on it. Behind him is a wooden wall with a gold-colored vertical paneling pattern. On the far left, there is a black speaker stand and a small table with a laptop and some papers.



Modern Data Architecture – Pipelining

JAVADAY2015



JAVA
DAY2015

<epam>

KLUXOFT
JUG.UA

ORACLE DATAART

Cogniance

GlobalLogic

SIGMA
Software

accenture

BRAINS

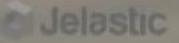
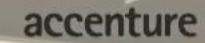
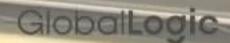
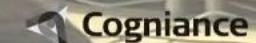
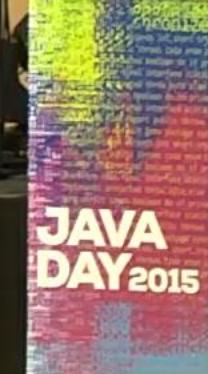
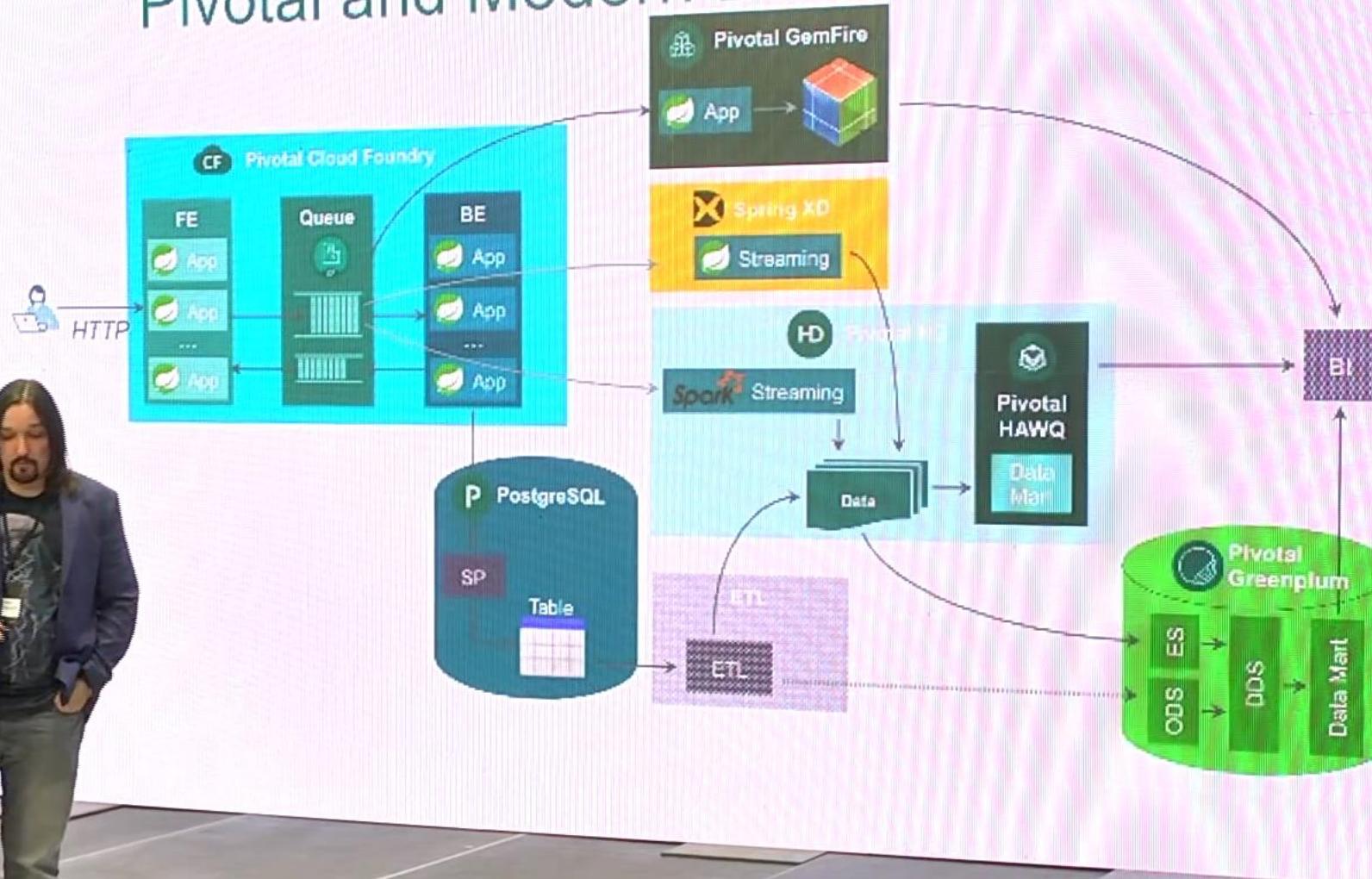
Jelastic

Playtika

INTRO PRO
SOFTWARE SERVICES AND PRODUCTS

Pivotal and Modern Data Architecture

JAVADAY2015



Chapter 8: Case Study

□ History of JD

□ Overview of used techniques in JD

□ Architecture

□ Software architect [软件架构师]

- 软件架构师是软件行业中一种新兴职业，工作职责是在一个软件项目开发过程中，将客户的需求转换为规范的开发计划及文本，并制定这个项目的总体架构，指导整个开发团队完成这个计划。主导系统全局分析设计和实施、负责软件构架和关键技术决策的人员

- 软件架构师一般都是具备计算机科学或软件工程的知识，**由程序员做起，然后再慢慢发展为架构师的。**
 - 在国内，很多大学目前还没有设立软件架构的学位课程，虽然IT业界对设计和架构的兴趣日渐高涨，但各学校还是无法在课程中增加相应的内容来体现这一趋势。从这个方面来说，学校教育已经远远落后于产业发展。

能力要求

<https://baike.baidu.com/item/软件架构师>

- 在技术全面、成熟练达、洞察力强、经验丰富，具备在缺乏完整信息、众多问题交织一团、模糊和矛盾的情况下，软件架构师能迅速抓住问题要害，并做出关键决定的能力、具备战略性和前瞻性思维能力，善于把握全局，能够在更高抽象级别上进行思考。主要包括如下：
- 1 对项目开发涉及的所有问题领域都有经验，包括彻底地理解项目需求，开展分析设计之类软件工程活动等
- 2 具备领导素质，以在各小组之间推进技术工作，并在项目压力下做出牢靠的关键决策；
- 3 拥有优秀的沟通能力，用以进行说服、鼓励和指导等活动，并赢得项目成员的信任；
- 4 以目标导向和主动的方式来不带任何感情色彩地关注项目结果，构架师应当是项目背后的技术推动力，而非构想者或梦想家（追求完美）；
- 5 精通构架设计的理论、实践和工具，并掌握多种参考构架、主要的可重用构架机制和模式（例如J2EE架构等）；
- 6 具备系统设计员的所有技能，但涉及面更广、抽象级别更高；活动确定用例或需求的优先级、进行构架分析、创建构架的概念验证原型、评估构架的概念验证原型的可行性、组织系统实施模型、描述系统分布结构、描述运行时刻构架、确定设计机制、确定设计元素、合并已有设计元素、构架文档、参考构架、分析模型、设计模型、实施模型、部署模型、构架概念验证原型、接口、事件、信号与协议等。



主要任务

<https://baike.baidu.com/item/软件架构师>

- 架构师的主要任务不是从事具体的软件程序的编写，而是从事更高层次的开发构架工作。他必须对开发技术非常了解，并且需要有良好的组织管理能力。可以这样说，一个架构师工作的好坏决定了整个软件开发项目的成败。
- 1 领导与协调整个项目中的技术活动（分析、设计和实施等）
- 2 推动主要的技术决策，并最终表达为软件构架
- 3 确定和文档化系统的相对构架而言意义重大的方面，包括系统的需求、设计、实施和部署等“视图”
- 4 确定设计元素的分组以及这些主要分组之间的接口
- 5 为技术决策提供规则，平衡各类涉众的不同关注点，化解技术风险，并保证相关决定被有效的传达和贯彻
- 6 理解、评价并接收系统需求
- 7 评价和确认软件架构的实现 专业技能



- 虽然大学要加强软件架构学课程的建设，但是，软件架构师的成长应该有一个实践的教育过程，并不是简单的学校的理论学习或者通过大型软件公司的认证就能成为合格的软件架构师。除了信息系统综合知识在学校学习外，软件架构师的大部分知识和经验将来自实际开发工作。根据软件架构师的任职条件，一名合格的软件架构师的成长应该经历8年以上的软件项目开发实际工作经验。一般需要经历程序员、软件设计师等阶段，然后再发展成为软件架构师。
- 当然，并不是每一位程序员经过8年后都可以成长为软件架构师的。一个软件工程师在充分掌握了软件架构师工作所必需的基本理论和技能后，如何得到和利用机会、如何利用所掌握的技能进行应用系统的合理架构、如何不断的抽象和总结自己的架构模式、如何深入行业成为能够胜任分析、架构为一体的精英人才，这就在于机遇、个人的努力和天赋了。

- 就目前来看，国内软件架构师的培养途径主要有两种方式，一种是大学（软件学院）教育方式，另一种是个人自我培养然后再进行相应的培训和认证。但是，不管哪种方式都有其不足之处。
 - 软件学院的培养方式能够系统的学习软件架构师必需的知识体系，但是，软件架构师不是简单的通过理论学习就能够培养出来的，软件学院的学生可能缺乏必要的设计、开发经验和相关的领域知识。尽管软件学院也强调给予学生实践的机会，但毕竟这种机会是有限的。有关“三分之一的师资来自企业”的规定，在部分软件学院中也没有得到真正落实，导致传授给学生的还是一些纯理论知识。
 - 自我培养方式的主要对象是具有一定年限的软件开发和设计人员，如Microsoft、IBM、Sun等公司的软件架构师认证对学员的基础并没有具体的要求，只要交纳规定的费用，然后进行几天的集中培训，通过考试就发给学员证书，甚至不需要考试就直接发放证书。这些开发人员在自我培养的过程中不一定能够系统的学习软件架构师的理论知识，他们只具有一定的开发和设计经验，仅仅经过几天的培训，是不太可能培养出合格的软件架构师的。而且，作为某个厂商的培训和认证，其最终目的是培育自己的市场，培养一批忠诚的用户，而不是为中国培养软件架构师。因此，也存在很大的问题和缺陷。

- 针对软件架构师在软件组织中的作用和其在国内的培养现状，有分析家认为有必要将软件架构师的教育、培训和认证作为发展民族软件产业的一个基本决策，制定详细的软件架构师培养方案。因此，提出以下一些关于软件架构师培养的基本方法和途径。
 - (1)确定软件架构师在软件组织中的职责和充当的角色，确定其相应的必须具备的知识体系，确定软件架构师的职业及其相关制度，制定软件架构师的培养目标和培养方案。
 - (2)坚持以大学教育为主（特别是各软件学院在这方面可以大施身手），以项目实践为辅的教育方针。大学可以聘请现有的软件架构师担任核心课程的讲师，通过学校教育，系统学习软件架构师所必需的知识体系；通过项目实践使其具有初步的软件开发和设计经验，逐步成长为一名合格的软件架构师。
 - (3)作为第2条的补充，聘请现有的软件架构师，个人认为适合以讲座的形式在学校开展。
 - (4)对国外一些大公司的软件架构师的培训和认证予以支持，但是在认证的过程中必须坚持符合中国实际情况的原则。例如，在认证考试之前对考生的知识体系进行系统的测试和评估，在通过认证后的适当时间内进行重新认证和继续教育。
 - (5)建立完善的软件架构师教育和认证制度，使得通过认证的人员能够在实际的软件开发中成为称职的和优秀的软件架构师。并通过此制度能够为国家培养出更多、更优秀的软件架构师，解决当前软件架构师急缺问题。