

百花齐放 锄其九九

Twitter的技术坎坷之路

Geekbang>

极客邦科技

整合全球最优质学习资源, 帮助技术人和企业成长
Growing Technicians, Growing Companies

InfoQ
LUCAS

专注中高端技术人员的
技术媒体



EGO EXTRA GEEKS' ORGANIZATION
NETWORKS

高端技术人员
学习型社交网络



StuQ
LUCAS

实践驱动的
IT职业学习和服务平台



GiT GEEKBANG
INTERNATIONAL
TRAINING
极客邦培训

一线专家驱动的
企业培训服务



旧金山 伦敦 北京 圣保罗 东京 纽约 上海
San Francisco London Beijing Sao Paulo Tokyo New York Shanghai

QCon

全球软件开发大会

2016年4月21-23日 | 北京·国际会议中心

主办方 **Geekbang**  **InfoQ**
极客邦科技

7折 优惠 (截至12月27日)
现在报名, 节省2040元/张, 团购享受更多优惠

www.qconbeijing.com



扫描获取更多大会信息

个人介绍

王天

2003年本科毕业于清华大学计算机系；

2005年硕士毕业于University of North Carolina Chapel Hill；

2005年7月加入Google，从事移动搜索、新闻搜索、搜索质量等工作；

2011年3月加入Twitter搜索部门，工作至今。

带领Twitter的搜索质量团队，改进实时搜索产品。



@wangtian

www.twitter.com/wangtian 

www.linkedin.com/in/wangtian 

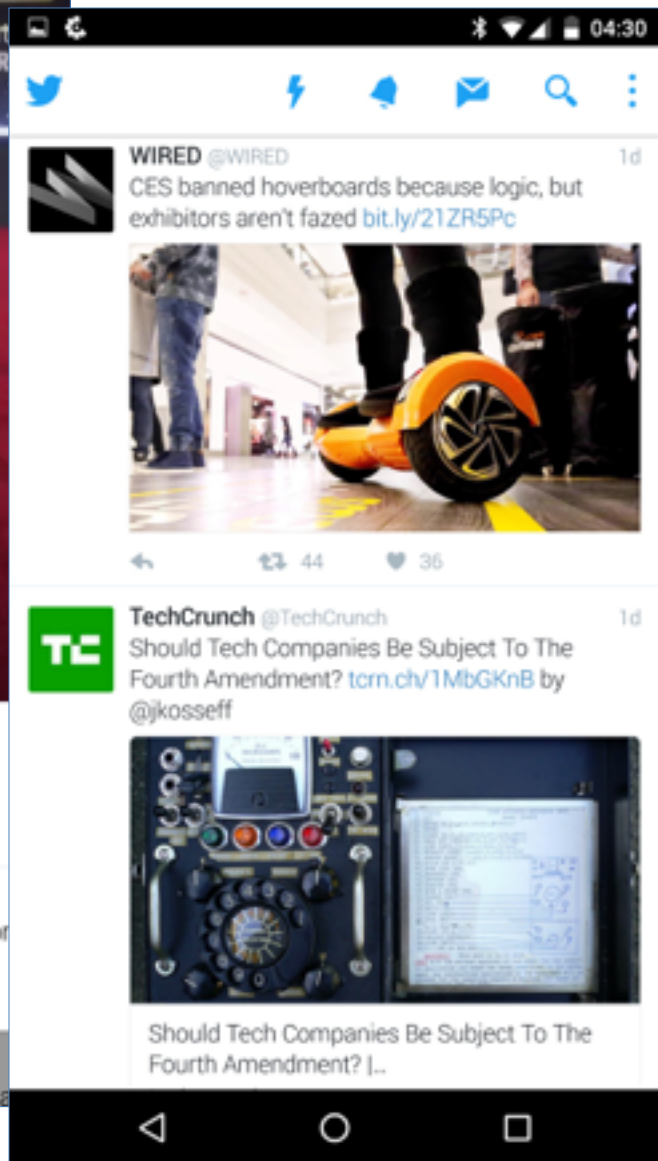
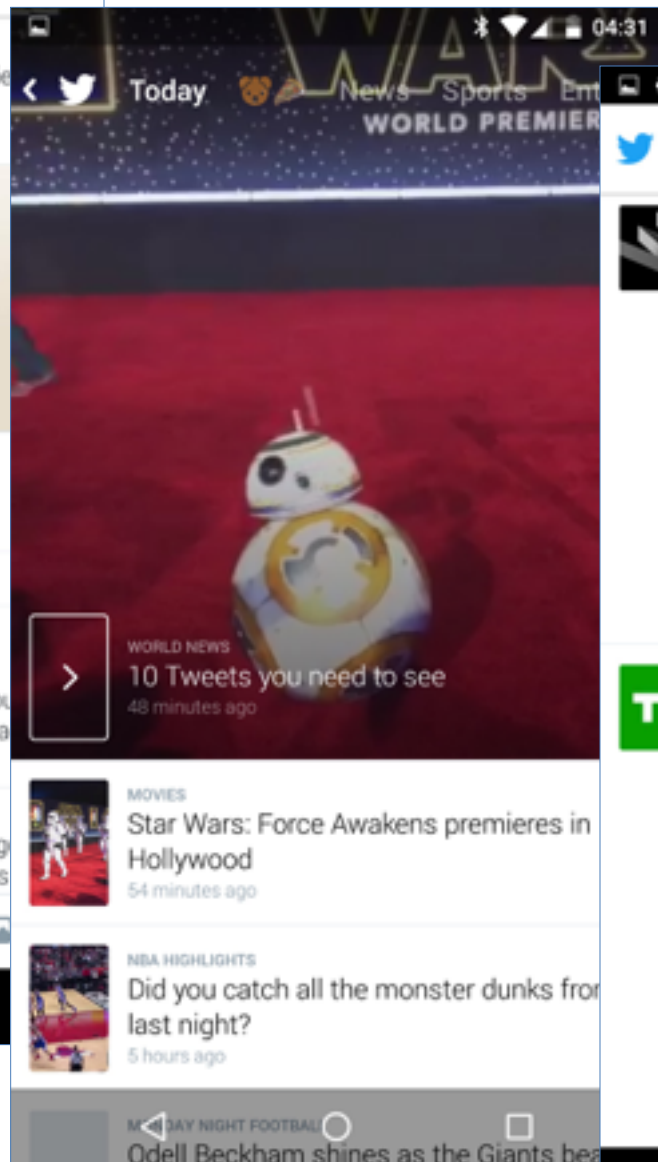
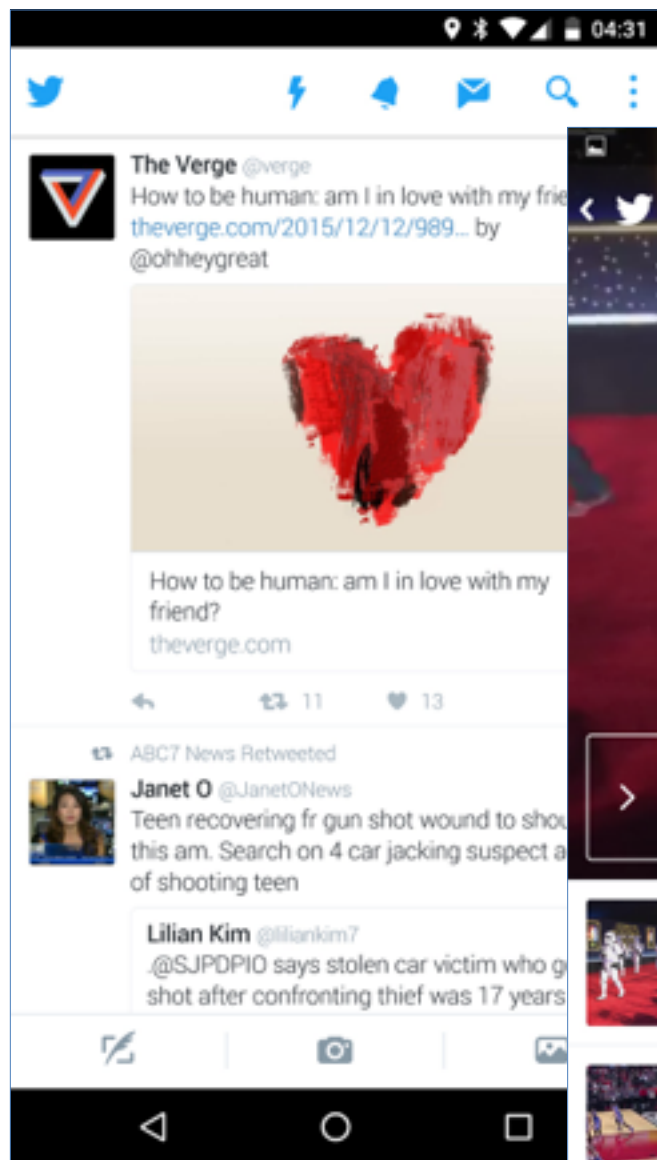
Twitter的基本数字

- 微博客始祖，2006年创始
- 3.2亿*月活跃登录用户
- 10亿*月活跃独立访问用户（包括网站嵌入推文）
- 80%*流量来自移动设备
- 79%*用户来自美国以外
- 4300*名员工，44%*为工程师
- 每日数亿条、每年逾2000亿条推文**



* 2015年9月数据

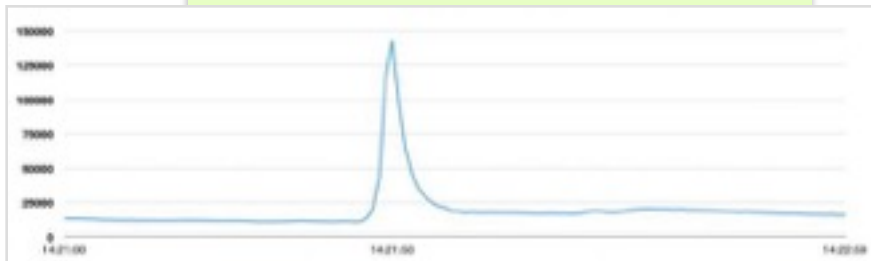
** 2013年



和世界息息相关的实时信息平台

可预测

奥运会开闭幕式
NBA决赛
NFL决赛
奥斯卡颁奖
日本《天空之城》重播 (2013.8)



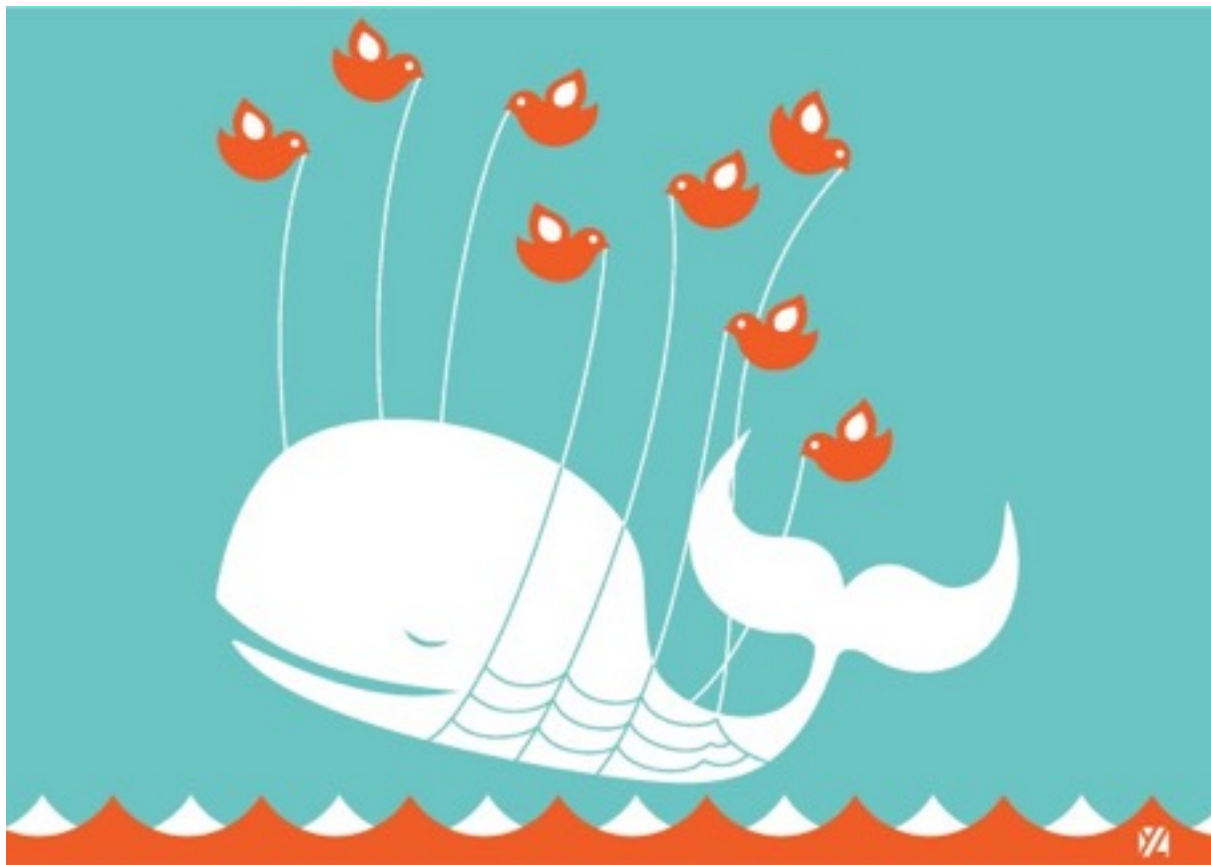
<https://blog.twitter.com/2013/new-tweets-per-second-record-and-how>

不可预测

日本海啸 (2011.3)
世界杯德国巴西半决赛 (2014.7)
奥斯卡Ellen自拍事件 (2014.3)
巴黎恐怖袭击
巴西音乐节的奇怪网站



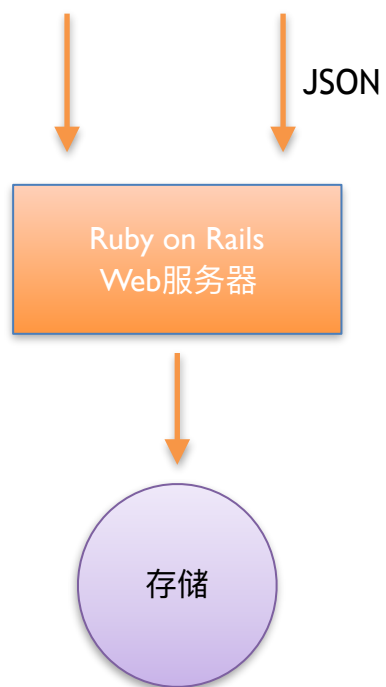
失败鲸 (Fail Whale)



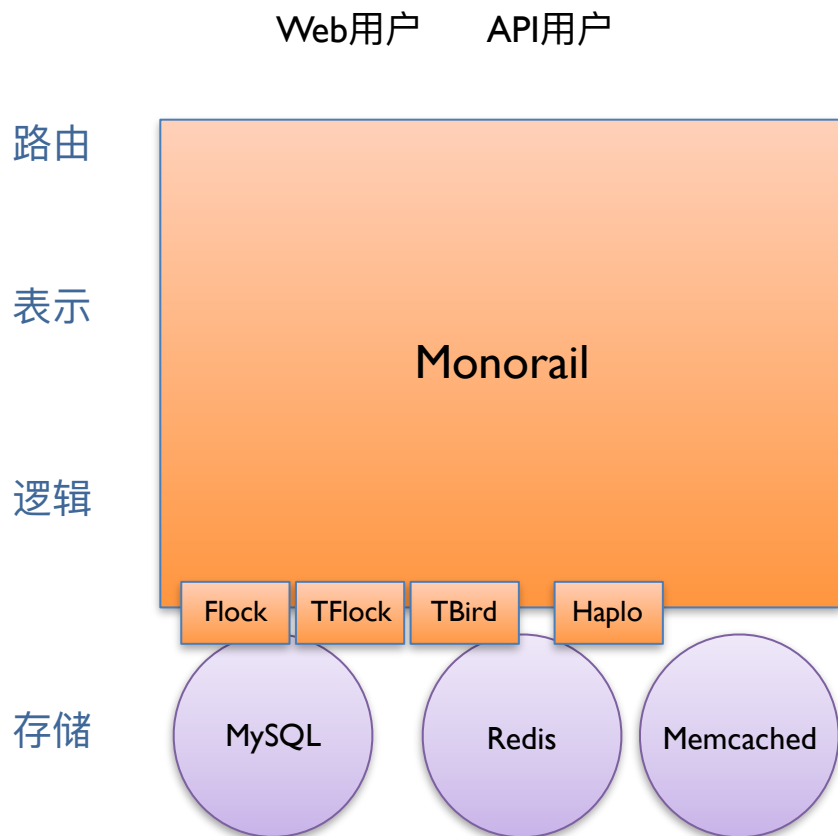
<http://www.yiyinglu.com/?portfolio=lifting-a-dreamer-aka-twitter-fail-whale>

远古

Web用户 API用户



古代

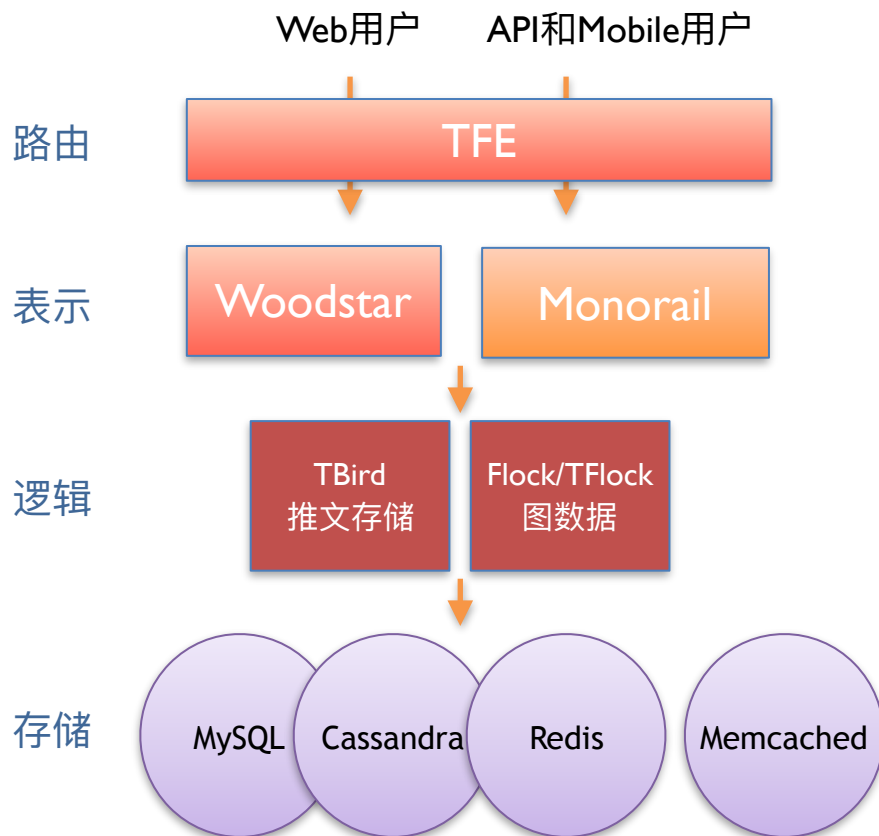


- 世界上最大的Ruby on Rails部署
- 所有逻辑都在Monorail中
- 超过200名工程师往里面check in代码
- 难以加入新功能，发布周期长
- 效率低下，延迟长，同步处理请求
- 单一数据库，热点明显
- 性能改善缓慢，增添机器的无底洞
- 没有很好挺过2010年南非世界杯的考验

技术债务累积迅速！

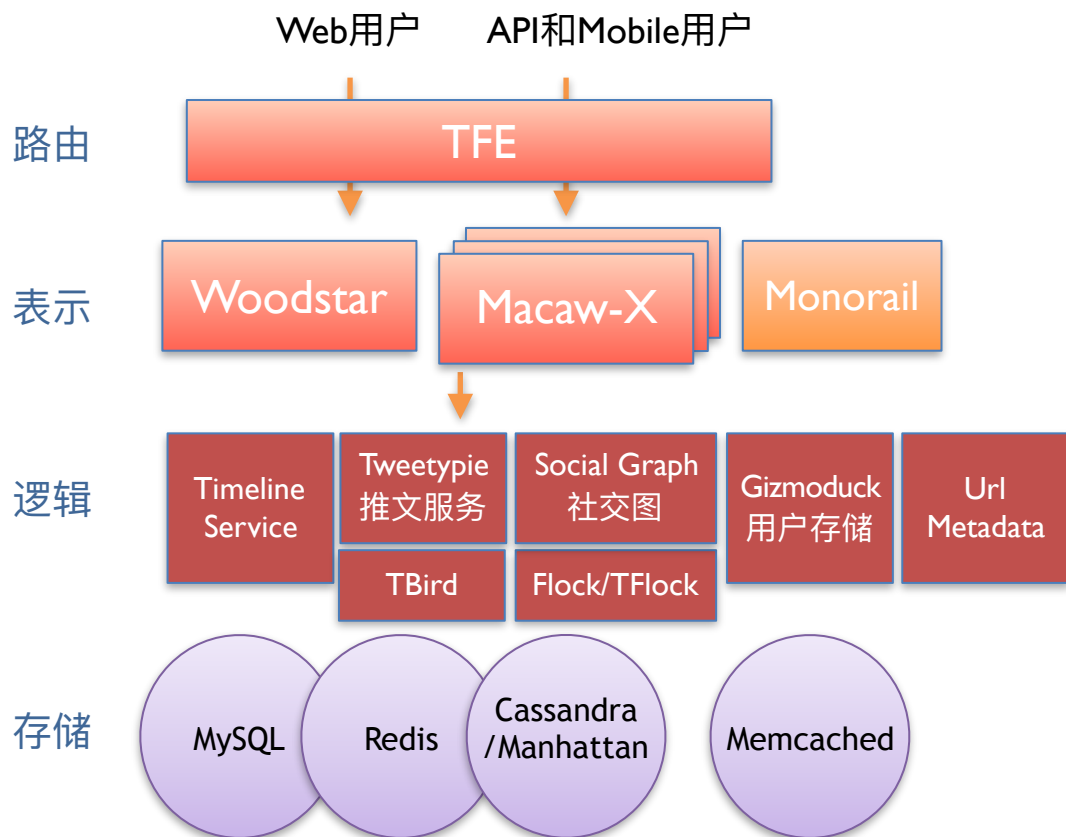
利率 = 在此上工作的人数

近代 (1)



- 减少耦合：分拆单一系统
- 提高效率：用Scala重写server
- 网络异步编程：开发Finagle
- 存储：创造较高级的数据服务

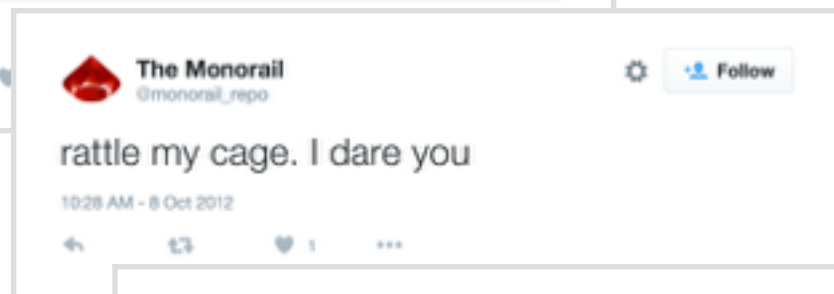
近代 (2)



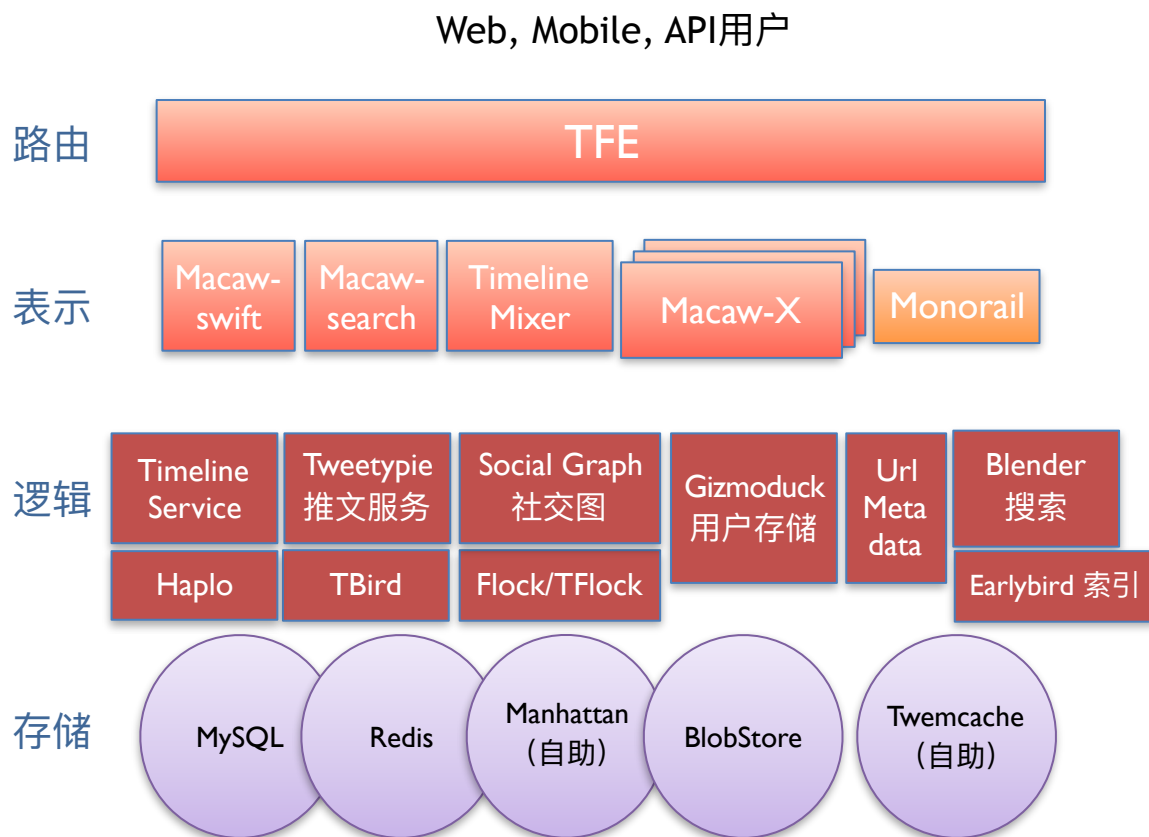
Monorail退休

Ruby vs. JVM

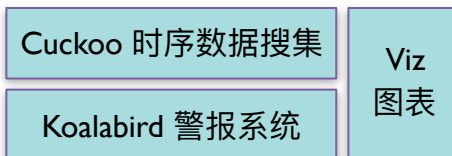
- 单机QPS处理能力从200~300提高到10000~20000，延迟减小到1/3
- 减少90%资源使用



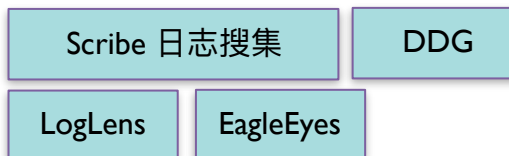
现代：产品系统和周边支持



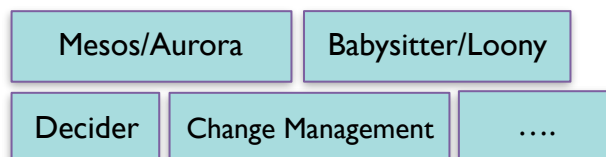
系统监测



数据分析



集群和运营



流处理

并行编程



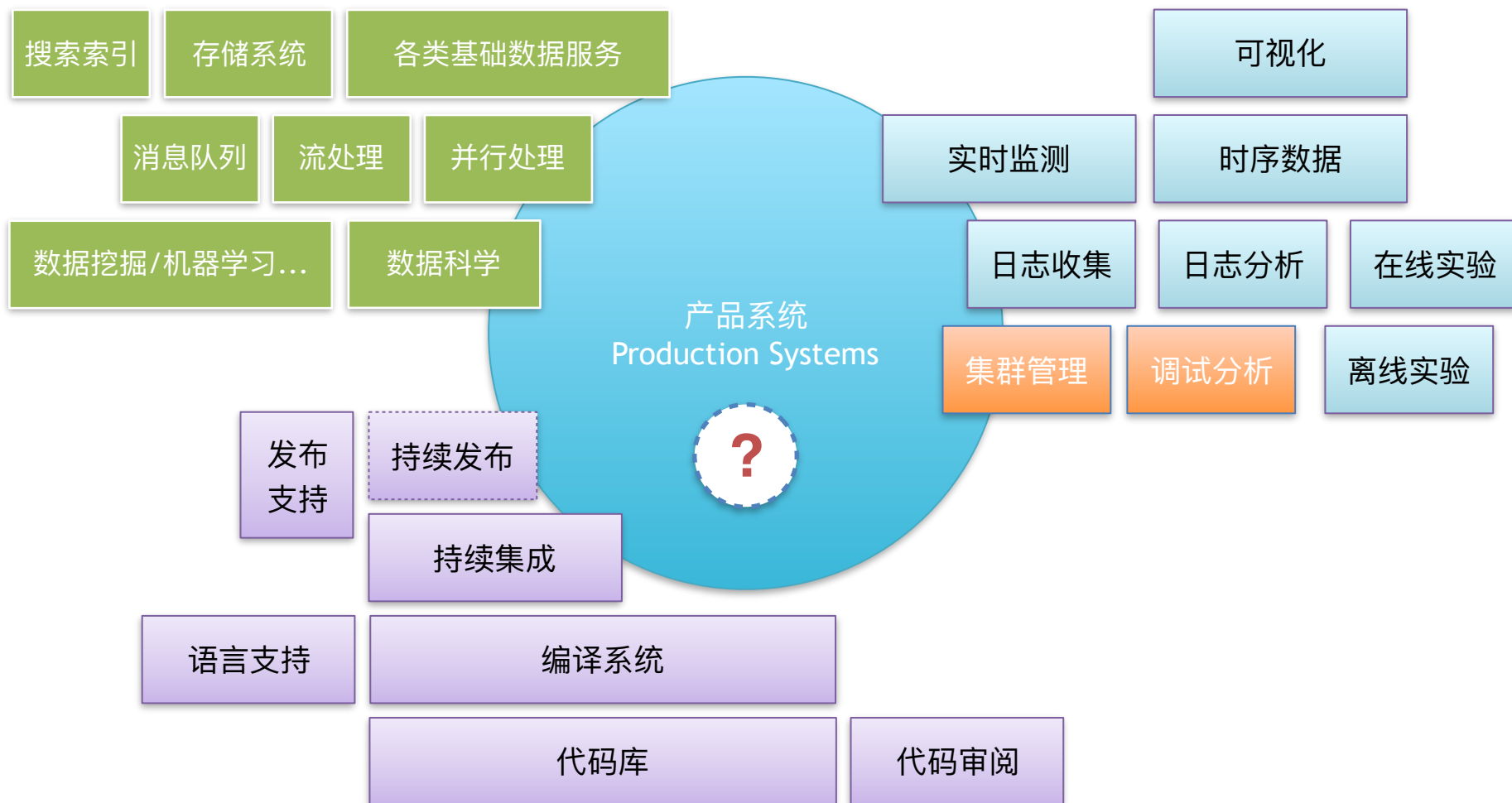
消息队列



RPC及跟踪调试

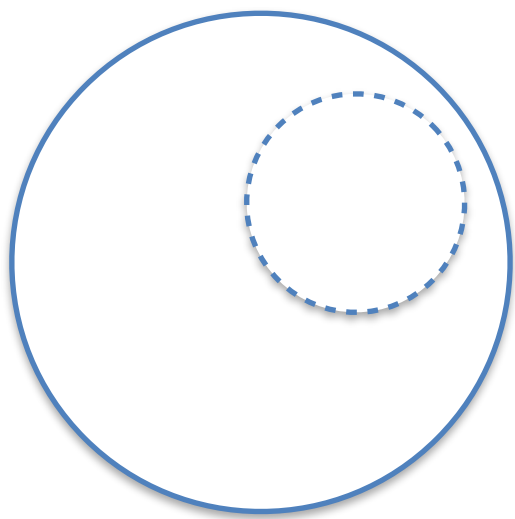


完整的工程生态



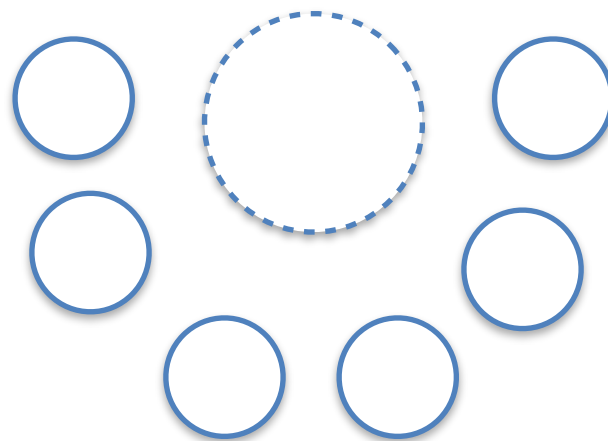
服务化构架

2009



知识无法压缩
关注不独立

2015

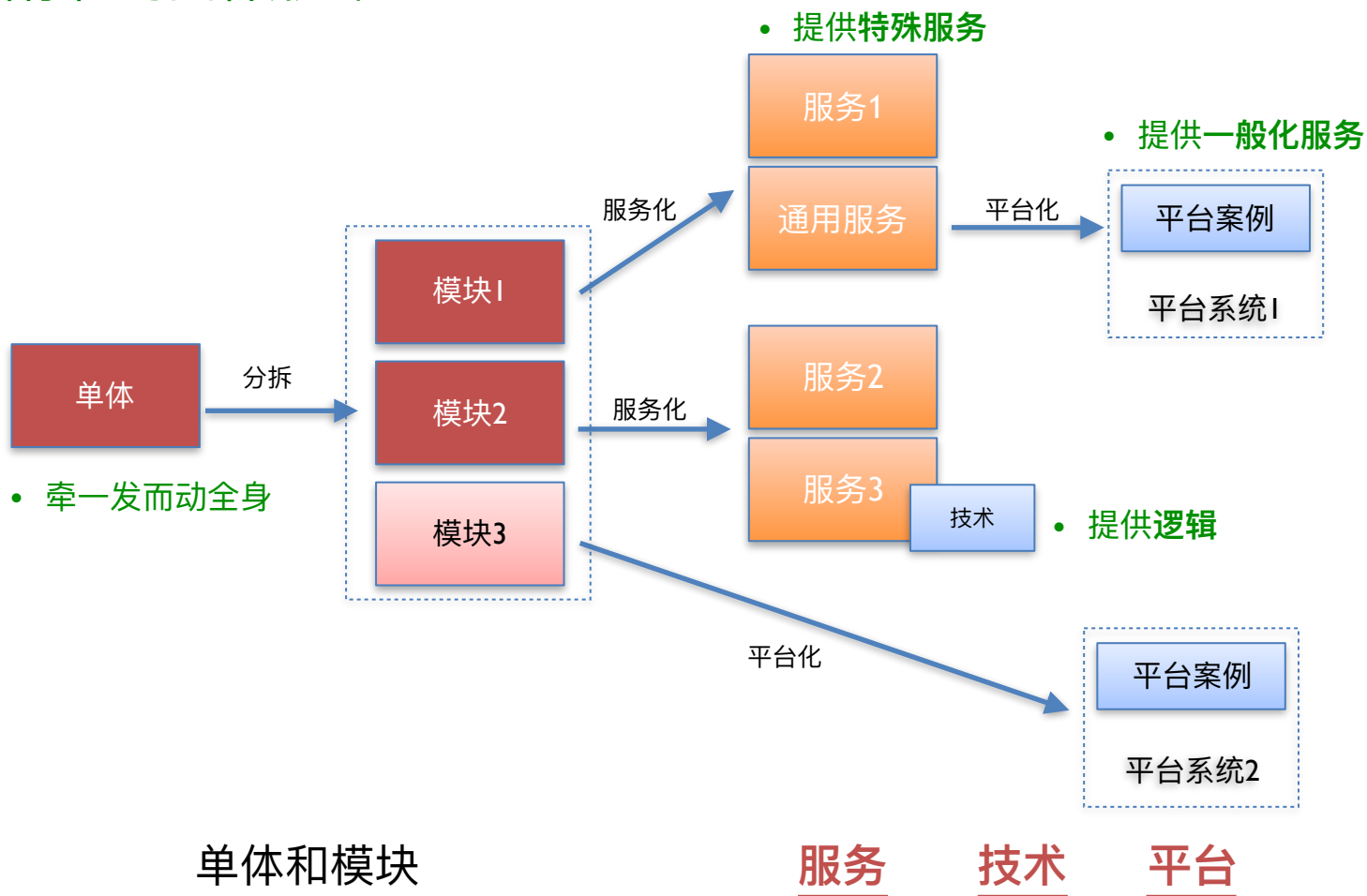


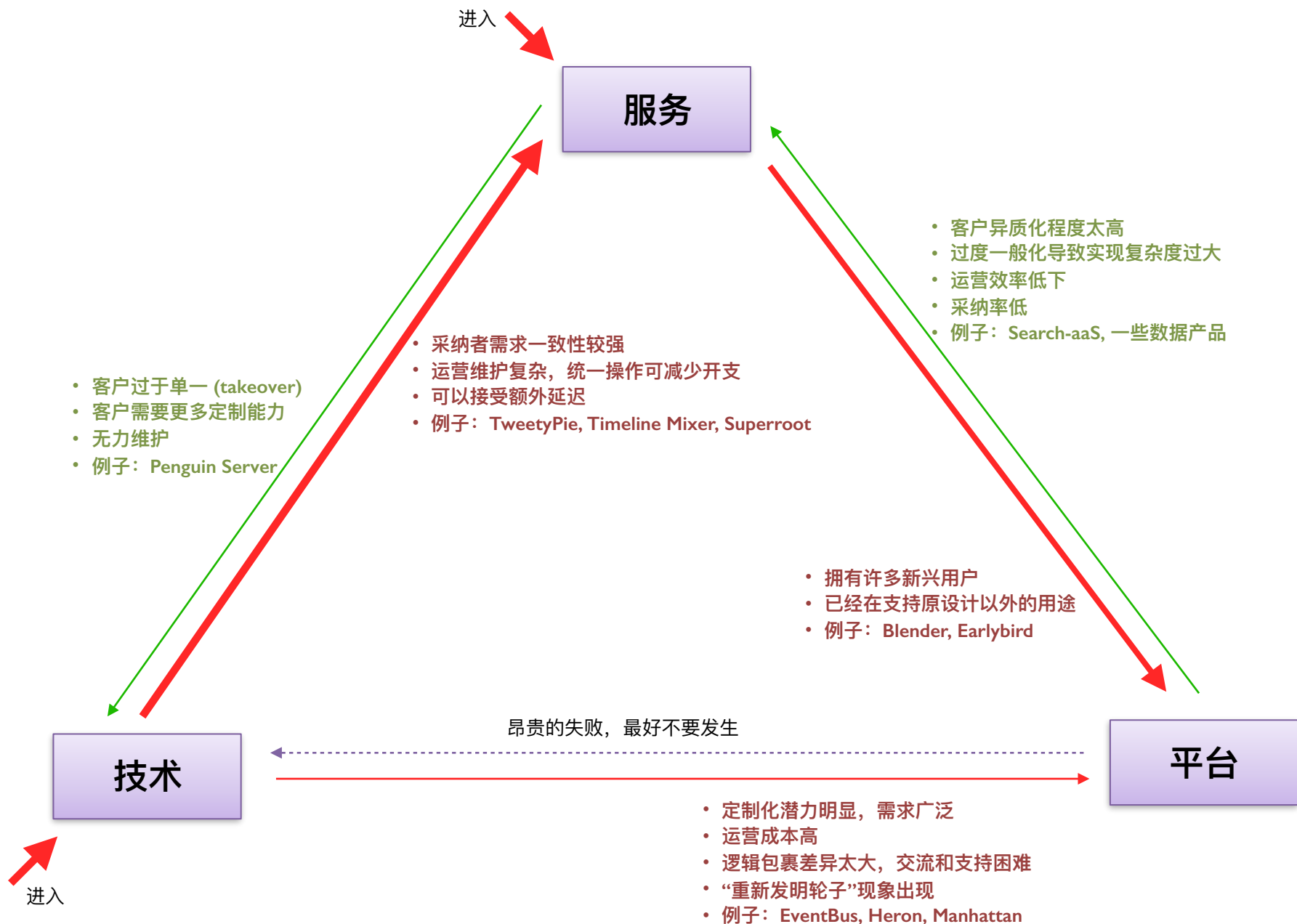
可集中精力关注业务逻辑
享受质量可预测的服务

Twitter的技术历史

2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
<ul style="list-style-type: none">• 服务器: Ruby on rails• 脚本: Ruby• 存储: MySQL• 搜索: MySQL		<ul style="list-style-type: none">• 服务器: Ruby, Java• 反向代理: TFE• 脚本: Ruby/Python• 集群: Babysitter/Mesos• 网络编程: Netty/Finagle• 并行计算: Pig• 流计算: Storm/自写• 存储: Redis/Cassandra• 队列: Kestrel• 缓存: Memcache• 搜索: Earlybird (Lucene)• 代码: 多代码库• 编译: Maven/Ants/Pants• Ruby on rails (退休中)			<ul style="list-style-type: none">• 服务器: Scala, Java, Ruby• 反向代理: TFE• 脚本: Ruby/Python• 集群: Babysitter/Mesos• 网络编程: Finagle• 并行计算: Scalding• 流计算: Storm/自写• 存储: Redis/Manhattan• 队列: Kestrel/Kafka/DistLog• 缓存: Memcache• 搜索: Earlybird/Blender• 跟踪: Zipkin• 代码: 多代码库• 编译: Pants为主• Ruby on rails (退休中)			<ul style="list-style-type: none">• 服务器: Scala, Java• 反向代理: TFE• 脚本: Python• 集群: Mesos/Aurora• 网络编程: Finagle• 并行计算: Scalding• 流计算: Summingbird/Heron• 存储: Manhattan(自助)/Redis/BlobStore• 队列: EventBus(自助)/DistLog/Kafka• 搜索: Earlybird/Roots/Blenders• 缓存: Twemcache(自助)• 监测: Cuckoo/Koalabird/Viz• 代码: (几乎)单一代码库• 编译: Pants为主• Ruby on Rails (彻底退休)	

逻辑存在的四种形式





当好一个服务生

- 原来是你同事的人现在是你的顾客

几条要领

- 用顾客需求驱动你的设计
 - 最简可行产品（MVP）
 - 不要实现既没有人需要也不能给你提供规划反馈的功能
 - 尽早实现效益
 - 部署之际已经能服务第一个客户
 - 考虑多顾客支持
 - 保证足够的灵活性
 - 注重客户体验
 - 好用的才会被采纳，被采纳的才能存活
- 用服务的语言来交流
 - SLA：明确服务期望
 - 思考“收费”模式
 - 创造市场和社区

你可能需要提供的东西

设计顾问
上手文档
示例代码
监测工具
客户支持
协调未来功能规划
...

持之以恒地推广

- 推广你的服务和实践
- 扩大它的用户群，增加采纳率
- 思考和类似服务共生和竞争的关系

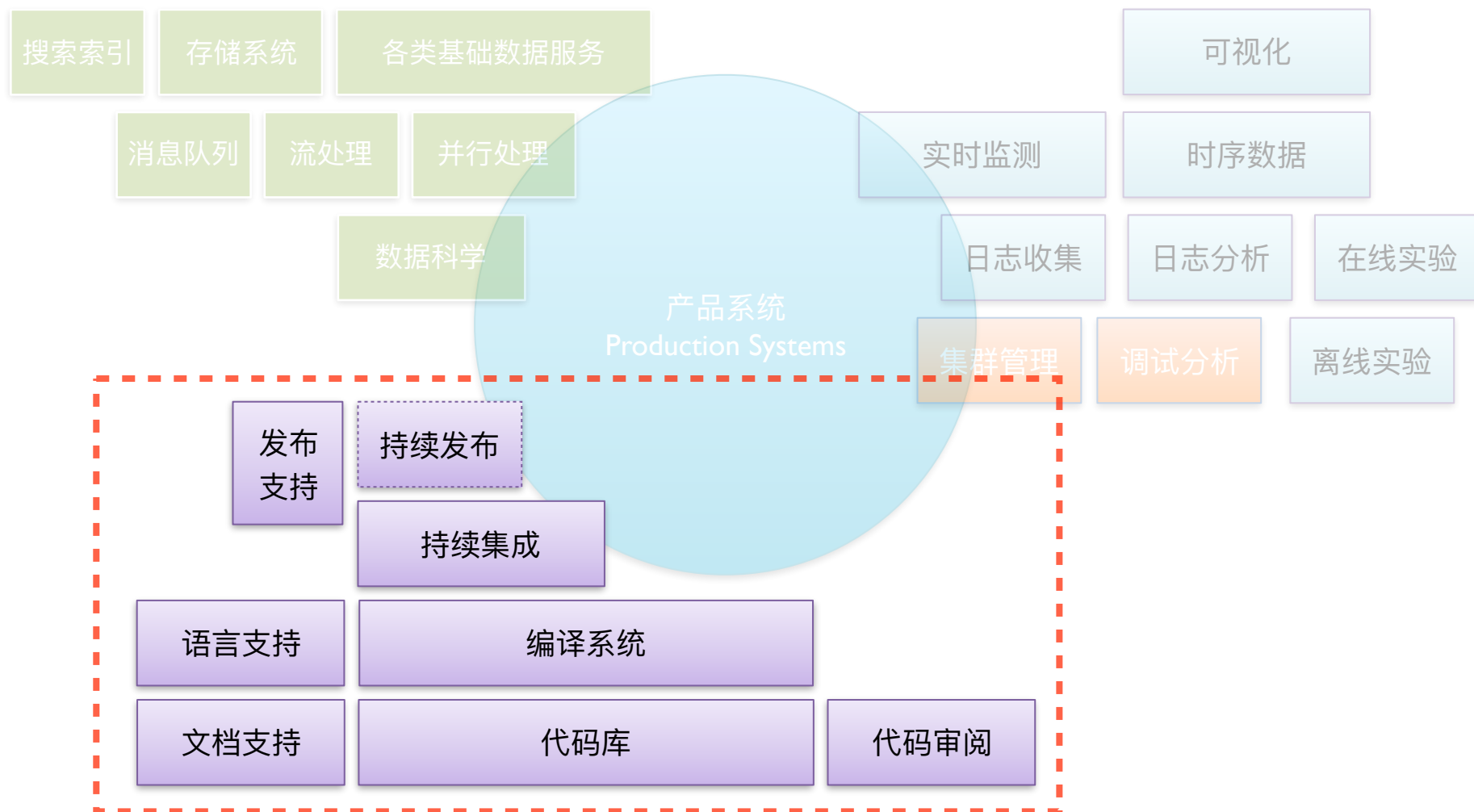
Evangelize!

设计底线

- 规范设计过程
 - 设计文档
 - 设计审核会议
- 确保符合当前最佳实践
- 有意识地提升工程质量底线
- 设立设计排查清单
- 充分讨论新技术引入的集成代价和支持代价
- 公开和协作
- 尽早引入利益方参与讨论
- 尽早思考产品化过程
- 设计导师：Design Shepherd

关于工程支持

另外一些幕后英雄



为什么工程支持对效率至关重要

假设一个工程师一年工作2000小时（250天）

效率提高比例	节省时间
1%	4.8分钟/天
2%	9.6分钟/天
5%	2.8小时/周
10%	2天/月

- 工程支持资源有限
- 把精力用在刀刃上
- 提升关键工具性能
- 增加关键工具适用范围
 - 锄草！

工程师的效率模型

工程师总数量

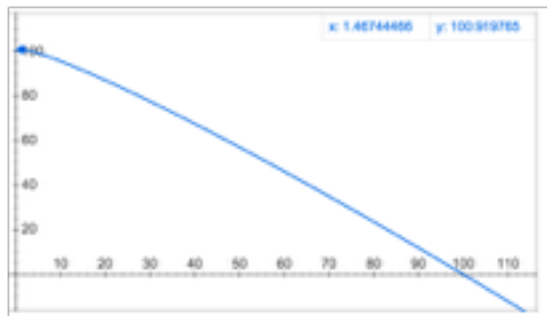
从事工程效率提升的人员数量

$$\text{效率} = (\text{eng} - \text{ee}) * (1 + b * \text{ee}^s)$$

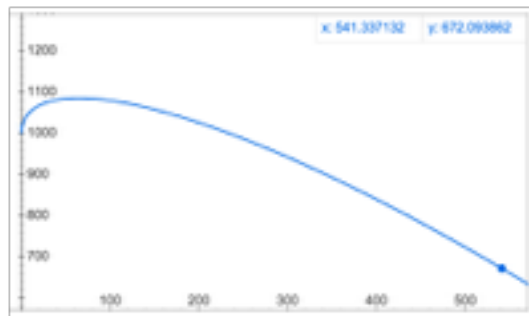
提升率

累积指数

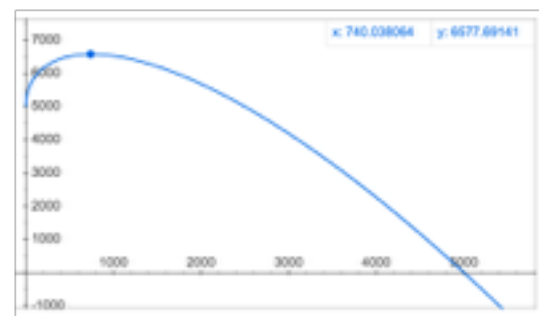
假设：b=2%，s=0.5



100个工程师: 1人



1000个工程师: 85人



5000个工程师: 740人

来源: <http://www.gigamonkeys.com/flowers/>

Twitter的语言支持演变

- Ruby

- Ruby
- Java (搜索)

- Ruby
- Java
- Scala
- Pig
- Python

- Ruby
- Java
- Scala (Ruby一样的)
- Scala (原教旨主义的)
- Scalding (并行计算)
- Pig
- Python
- C/C++ (特殊开发)

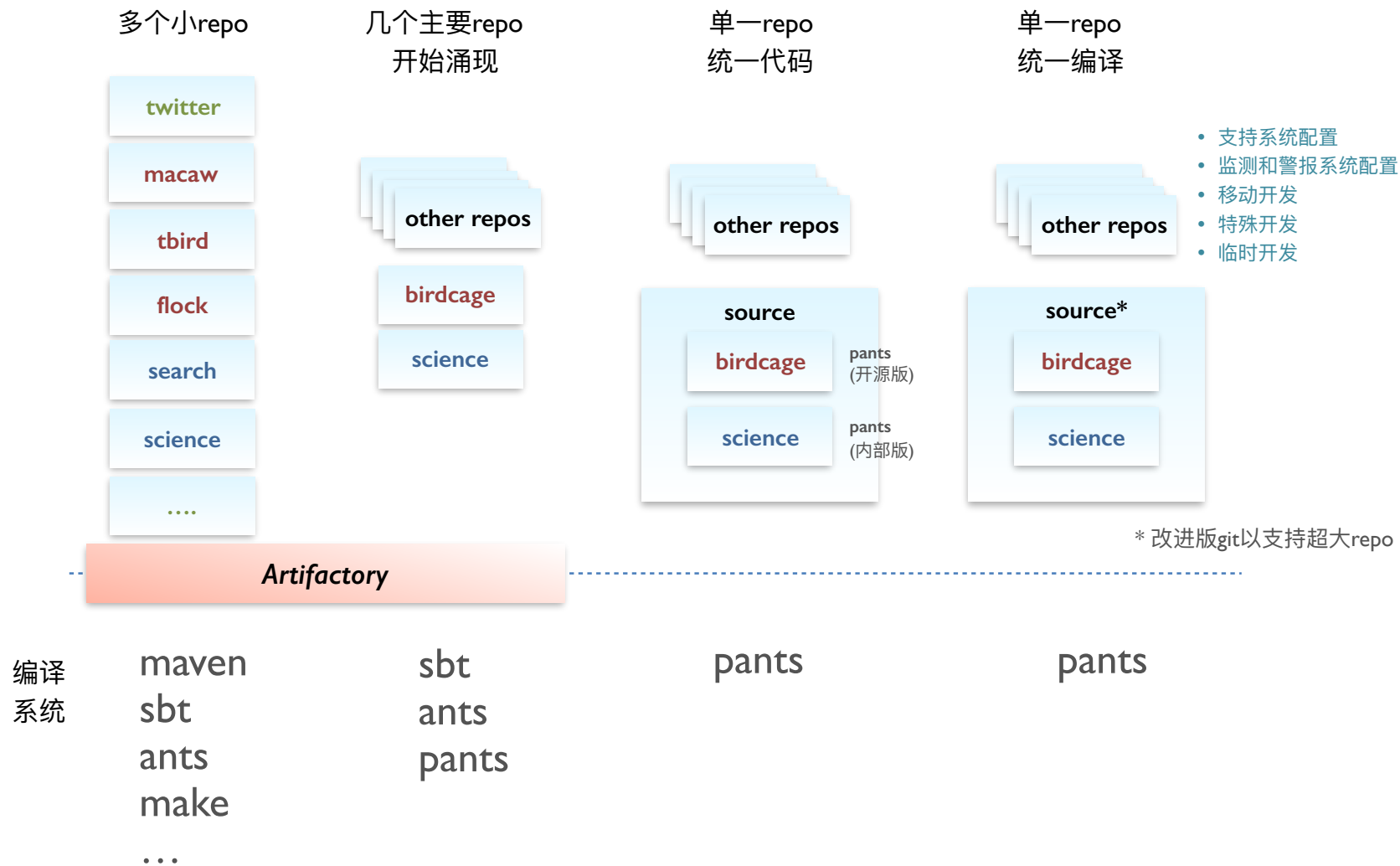
- Java (搜索、广告、Android)
- Scala (后台)
- Scalding (并行计算)
- Python (脚本)
- C/C++ (特殊开发)

Thrift

Thrift for
Finagle

Thrift
Scrooge

Twitter的代码库和编译系统演进



质量和运营底线

- 设计审阅、代码审阅
- 软件测试
 - 单元测试
 - 集成测试
 - 试运营测试
 - 压力测试
- 系统可观测
 - 实时观测
 - 事后分析
 - 数据驱动决策的根本
- 可运行时调整
- 过程可跟踪
- 后事有料理

其他工具

- ReviewBoard：代码审核工具
 - 经过扩展以匹配公司代码审阅流程
- JIRA：任务规划和追踪
 - 各团队自行选择任务产生、分配和规划方式
- Confluence：公司内部Wiki
 - 维护团队文档、内部资料，指南等
- HipChat：聊天室
- DocBird：自开发和代码库集成的技术文档系统
- Google Docs
 - 协同编辑和审阅文档，共享文档、表格、幻灯片
- Google Calendar
 - 日历安排和协调

开源

kestrrel twurl
flockdb
elephant-bird commons twitter-server diffy
twitter-text twitter4j
Finagle Scrooge Pants twemcache
bijection Twemproxy gizzard mysos
github.com/twitter
iago ambrose Scalding watchman
Summingbird libcrunch
hraven util snowflake jvmgcprof
cassovary typeahead.js





Following

software engineering



RETWEETS

61

LIKES

51



4:18 PM - 19 Nov 2013



Thanks!

