

ArchSummit全球架构师峰会 北京站2015

DT时代下[个推3.0]遵循的四个法则

叶新江 CTO@个推

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
UEUE

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
UEUE

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

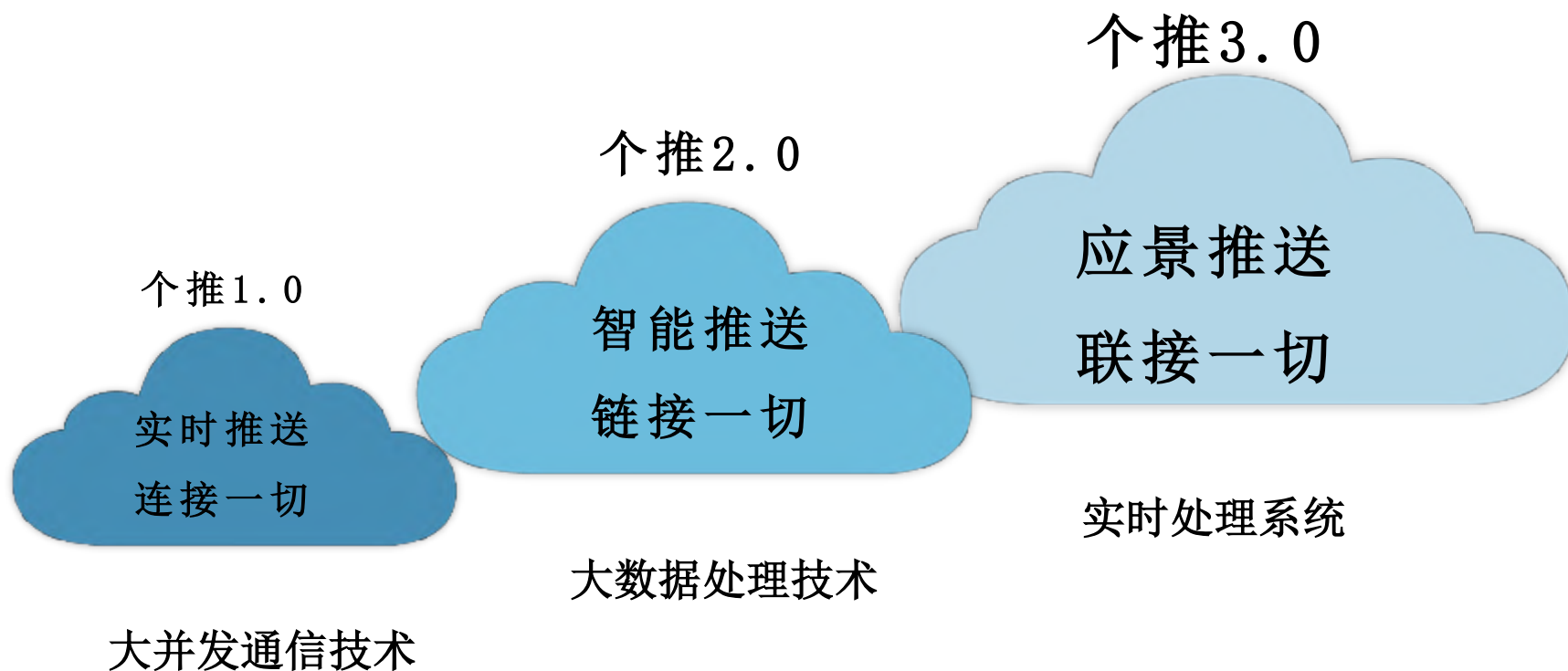
7折 优惠 (截至12月27日)
现在报名, 节省2040元/张, 团购享受更多优惠

www.qconbeijing.com



扫描获取更多大会信息

产品发展路径



个推3.0所遵循的四个法则

- 数据是决策的基本依据
- 数据越热越有价值
- 近似优于精确
- 永远的墨菲定律



数据是几乎所有决策的基础



我们所遵循的四个法则（2）- 数据决策

例子1：

是否引入新的处理工艺？

- 解决什么问题？
- 当前存在什么问题？
- 在当前业务场景下性能数据如何？
- 出现问题时，解决的成本如何？
- 是否是为了逃避前一个问题？



我们所遵循的四个法则（2）- 数据决策

例子2：

数亿客户端情况下，如何定位问题？

譬如：

有的手机上正常，有的不正常.....

有的区域正常，有的不正常.....

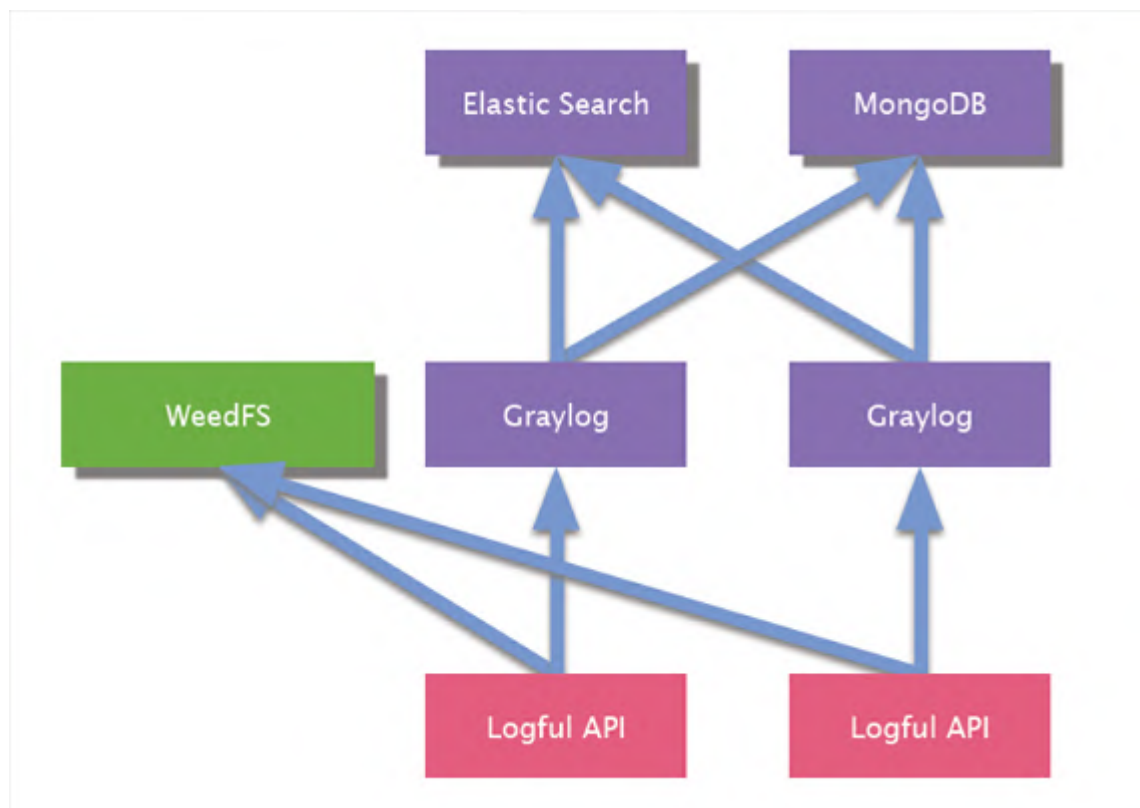
有的版本正常，有的不正常.....

对策：数据抽样收集、集中分析



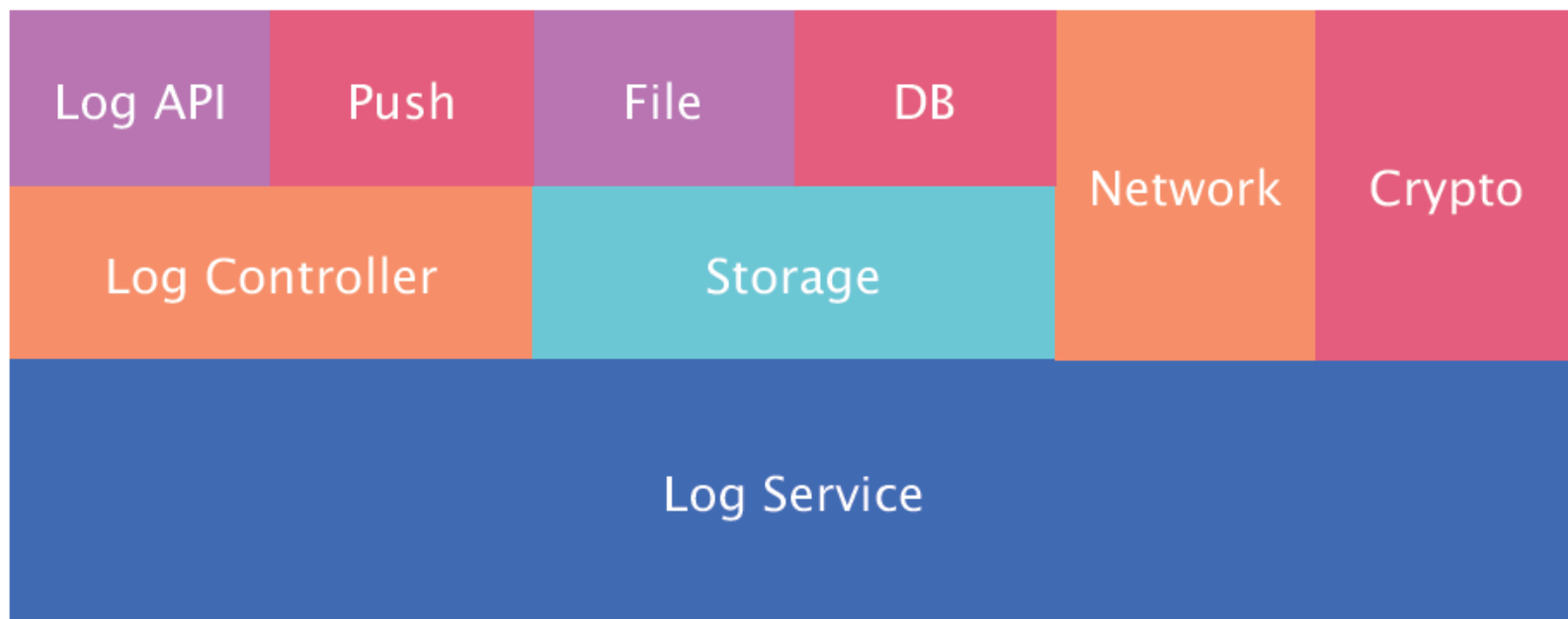
我们所遵循的四个法则（2）- 数据决策

Logful (<https://github.com/logful>)



我们所遵循的四个法则（2）- 数据决策

Logful SDK 架构



例子3：

性能测试和验证

性能测试环境可以测试到什么程度？

在生产环境下是否可以有足够数据来反过来印证？

对策：性能监控，及时发现局部问题，立即着手解决



数据越热越有价值



我们所遵循的四个法则（2）- 数据价值

数据的热度划分：

- **冷数据**

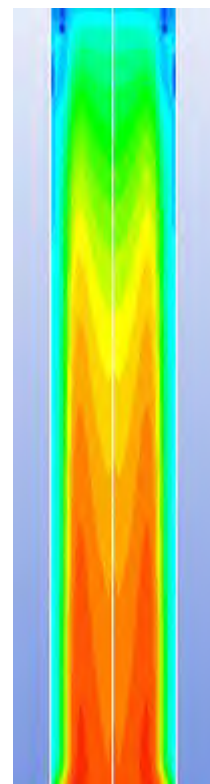
较长时间之前的状态数据、用户画像数据

- **温数据**

非即时的状态和行为数据

- **热数据**

即时的位置状态、交易、浏览行为



我们所遵循的四个法则（2）- 数据价值

例子1：

- 应景推送

利用冷数据 + 热数据进行实时处理

方案比较：

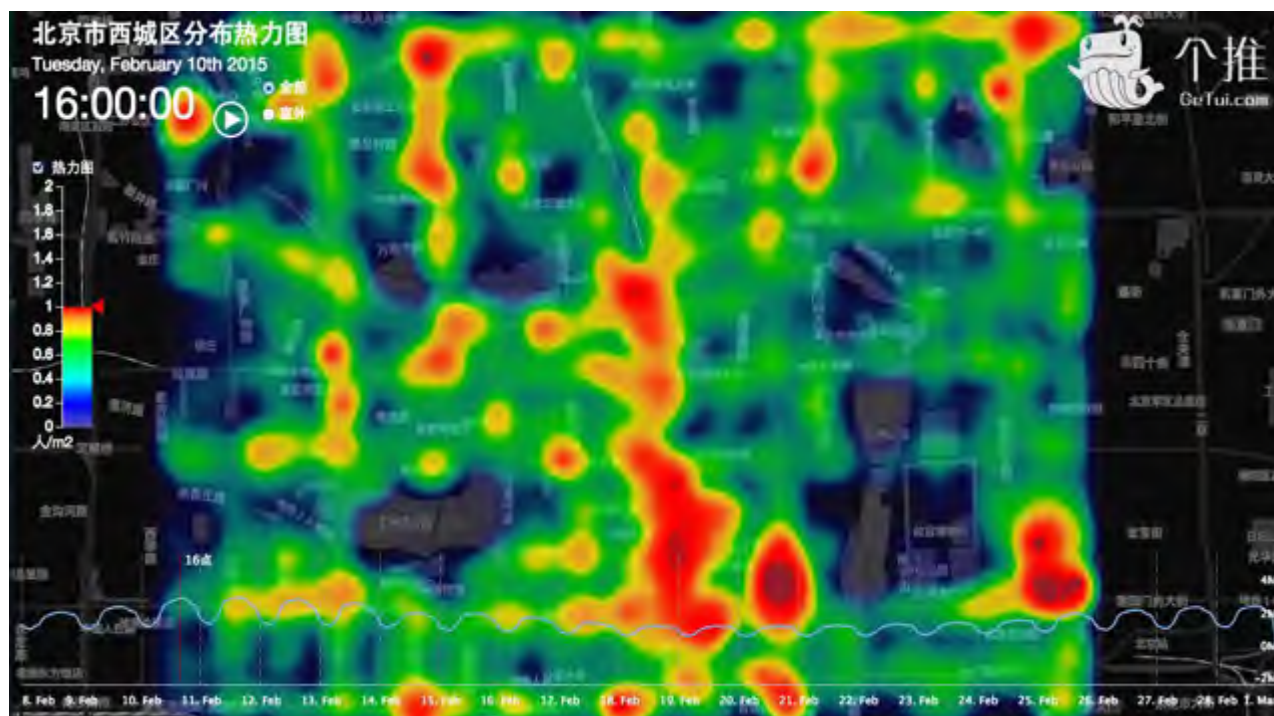
1. 客户端处理
2. 服务端处理



我们所遵循的四个法则（2）- 数据价值

例子2：

- 个图（人口热力图）



我们所遵循的四个法则（2）- 数据价值

例子2：

- 个图（人口热力图）

准实时要求高，并且能动态表现

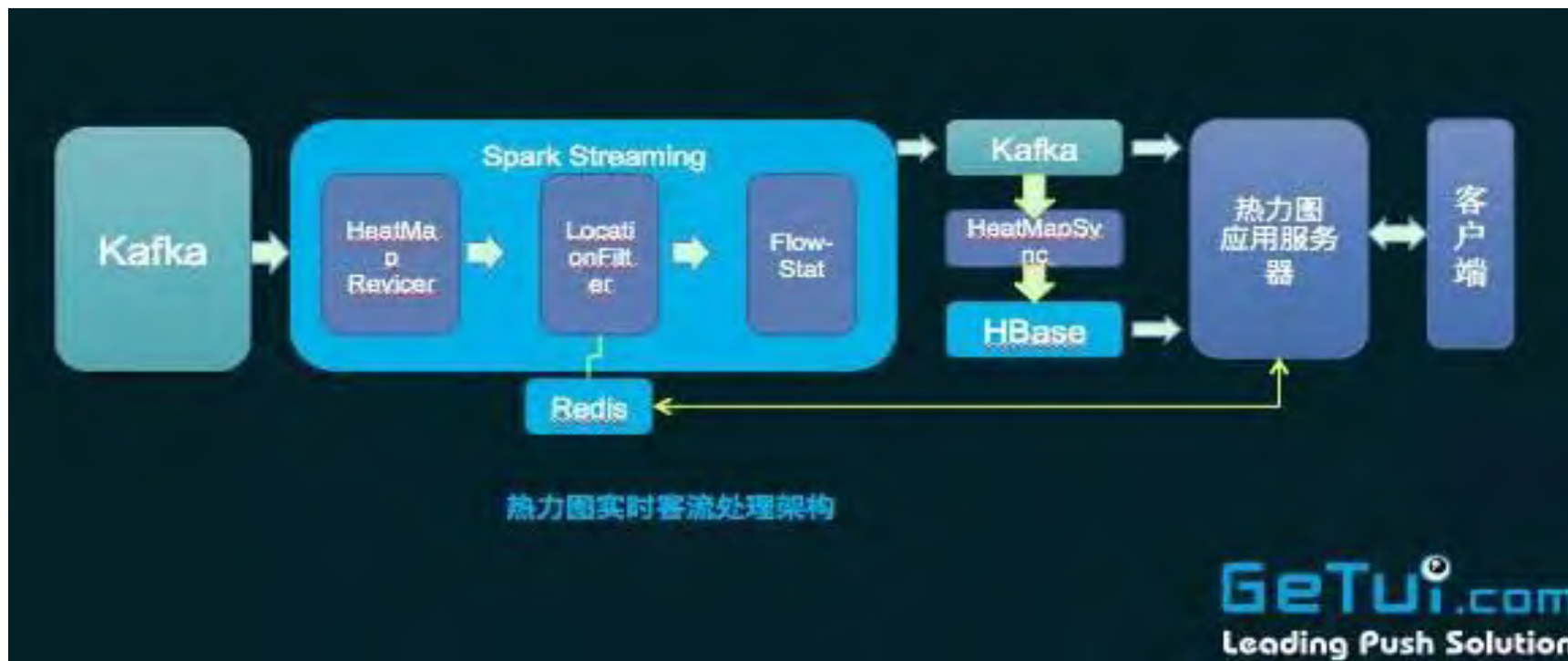
对策：

分级、分区域、分阶段



我们所遵循的四个法则（2）- 数据价值

个图架构：



近似优于精确



我们所遵循的四个法则（3）- 近似优于精确

考虑以下场景：

需要统计流数据中独立元素的个数？

要求：实时，至少是准实时



我们所遵循的四个法则（3）- 近似优于精确

考虑因素：

- **数据流速度很快**

意味着无法使用二级存储

- **数据规模巨大**

意味着要么使用超大内存的设备, 多大算

是个头？

要么多个设备分而治之运算



我们所遵循的四个法则（3）- 近似优于精确

一般化的结论：

- **获得问题的近似解比精确解要高效得多**
- **一系列与哈希相关的技术被证明十分有用**



Flajolet-Martin算法：

一个简单直观的基数估算方法

- 1、随机生成n个服从均匀分布的数字
- 2、随便重复其中一些数字，重复的数字和重复次数都不确定
- 3、打乱这些数字的顺序，得到一个数据集



永远的墨菲定律



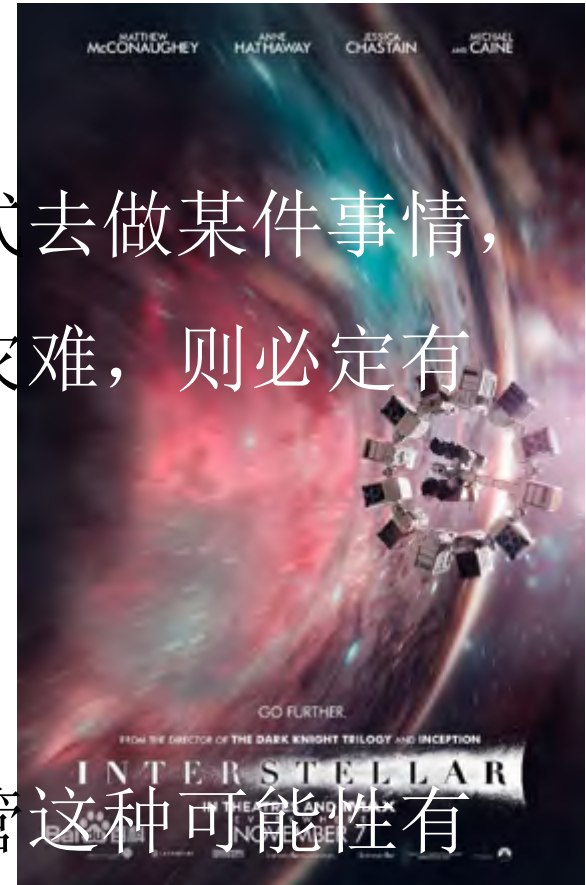
我们所遵循的四个法则（4）- 永远的墨菲定律

原句：

如果有两种或两种以上的方式去做某件事情，而其中一种选择方式将导致灾难，则必定有人会做出这种选择。

通俗点(Finagle's Law)：

如果事情有变坏的可能，不管这种可能性有多小，它总会发生。



我们所遵循的四个法则（4）- 永远的墨菲定律

例子：

不管内网如何快，总有断的时候。

- Redis 2.8 earlier 版本在网络闪断情况下的问题

流量并不一定按照我们的预期来到

- 流控未执行到位

还有：攻击、域名劫持、设备断电。。。。。

对策：

异常情况分析 + 预案设定 + 沙盘推演 + 模拟操作



个推，最专业的推送技术服务商

真诚期待与您合作



微信关注个推



个推官网注册

- 技术支持：kegf@getui.com
- 商务合作：zhangyao@getui.com
- 联系电话：0571-87971893

Thanks!

