



数据驱动研发效能： 指标、系统与实践

任晶磊 思码逸CEO

在 *DevData Talks* 开放务实地聊聊研发效能



分享合集



研效交流群

💡 每月直播 干货满满

行业 KOL + 企业研发效能负责人 + 资深效能顾问
从效能建设路径到不同场景应用，分享前沿观点与先进实践

💡 开放探讨 共同成长

直播 Q&A + 交流社群
直播嘉宾面对面答疑解惑，交流群随时随地探讨效能话题

关于思码逸

作为 DevData Talks 的发起方，思码逸专注为企业研发团队提供效能数据分析，呈现效率、质量、人才发展等多视角数据洞察，辅助团队决策与工程改进。

思码逸已服务 腾讯、美团、滴滴出行、中国平安、泰康保险、戴尔 EMC 等众多行业标杆客户。

讲师简介



任晶磊
思码逸 CEO

“

- 清华大学计算机系博士，前微软亚洲研究院研究员，斯坦福大学、卡内基梅隆大学访问学者
- 《软件研发效能度量规范》标准核心起草专家
- 在 FSE、OSDI 等顶级国际学术会议上发表多篇论文
- 参与过微软下一代服务器架构的设计与实施，获得 4 项美国发明专利
- Apache DevLake、CNCF DevStream 开源核心贡献者

”

摘要

- “低增长”新常态，对程序员意味着什么？
- 研发效能的驱动力
- 数据驱动研发效能的四个阶段
- 数据驱动研发效能的实现和案例



“低增长”新常态意味着什么



你们裁员了吗？

“低增长”周期——低速度、高质量

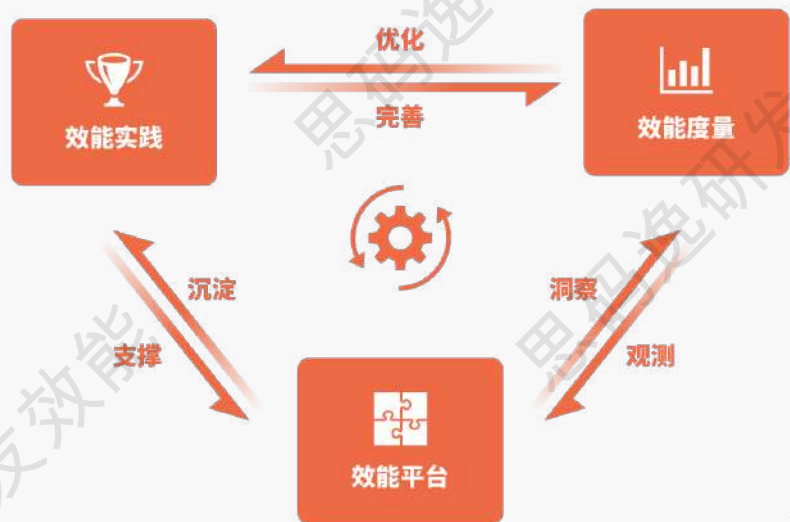
研发效能升级是大势所趋

用好数据，规划健康的软件工程

为什么“趋势”就一定不可逆转的呢？
因为提高了效率。

——刘润

研发效能的驱动力



研发效能的黄金三角 @ 张乐, 2021.07

流程、工具、数据是软件研发效能的三个主要驱动力

“凯撒的归凯撒，上帝的归上帝”

- 把流程和工具的选择权交回给一线团队
- 把工程一致性构建在数据上



Ready?

- 管理先于数据
- 数据辅助人
- 度量是统计性的

Go!

数据驱动研发效能的四个阶段

阶段	特征	陷阱	最佳实践
数据积累	有意识积累研发数据，将其视为资产和信息源	追求指标多而全，执着于个别指标的高低	利用开源系统和指标模型，快速跨越数据治理的鸿沟
数据夯实	维护数据健康度，建立高质量数据源	缺乏可靠的数据基座，在指标上虚假繁荣	采用代码分析数据，并与其他领域数据交叉关联
自主分析	利用BI或可视化工具，从数据中解读信息	围绕数据而非目标；缺乏适合研发模型的下钻，只能满足浅层信息需要	采用 GQM（目标-问题-指标）方法，通过合理下探推动 MARI（度量-分析-回顾-改进）循环
系统洞察	以目标为导向，基于健康的数据，合理分析和洞察	多样而有特异性的分析，强依赖分析师个人，无法在组织中规模化应用	参考规范化 GQM 模板，自主实现洞察和 MARI 循环；进一步复用和丰富模板，沉淀知识

数据驱动研发效能的实践和案例 I

阶段	特征	陷阱	最佳实践
数据积累	有意识积累研发数据，将其视为资产和信息源	追求指标多而全，执着于个别指标的高低	利用开源系统和指标模型，快速跨越数据治理的鸿沟

是我们用 gitlab 数据然后我们一个同事说 devlake 很不错

同事在小米待了 9 年

我 manager 是公司技术主管吧

他提出的一些想法

你们公司直接就有这些产品

你们这个产品很牛啊



<https://devlake.apache.org/>

研发大数据平台



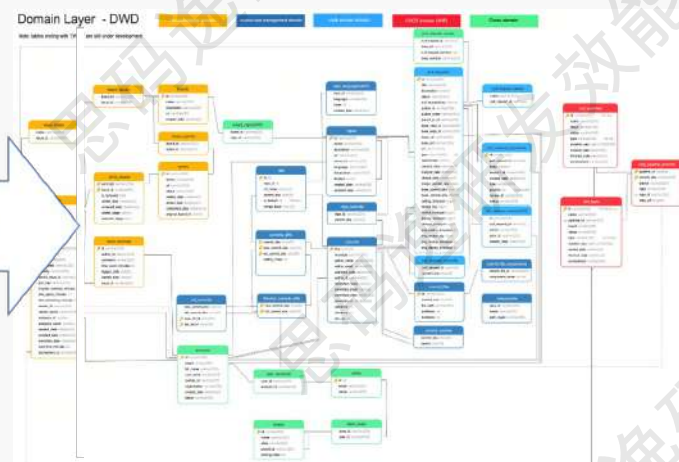
Fork 275 | Starred 1.8k



- 需求管理模型
- 代码管理模型
- 测试用例模型



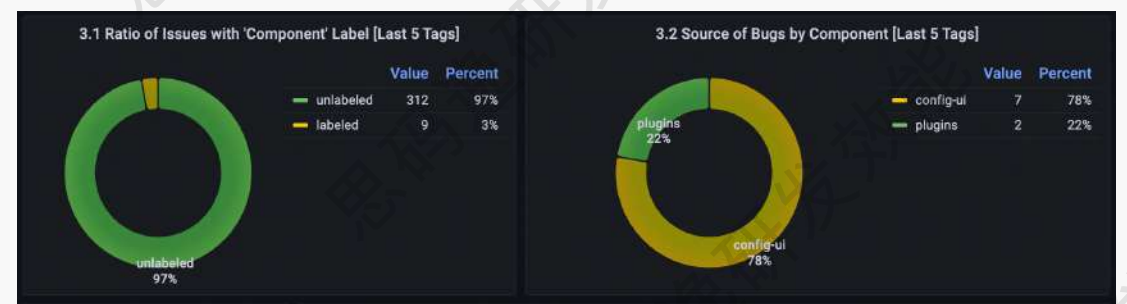
- 持续集成模型
- 部署环境模型
- 通用领域模型



数据驱动研发效能的实践和案例 I

ClickHouse

- <https://devlake.apache.org/livedemo/DataSources/GitHub/>
- <https://devlake.apache.org/livedemo/QAEngineers/GitHubReleaseQualityAndContributionAnalysis/>



2.7 Work Distribution

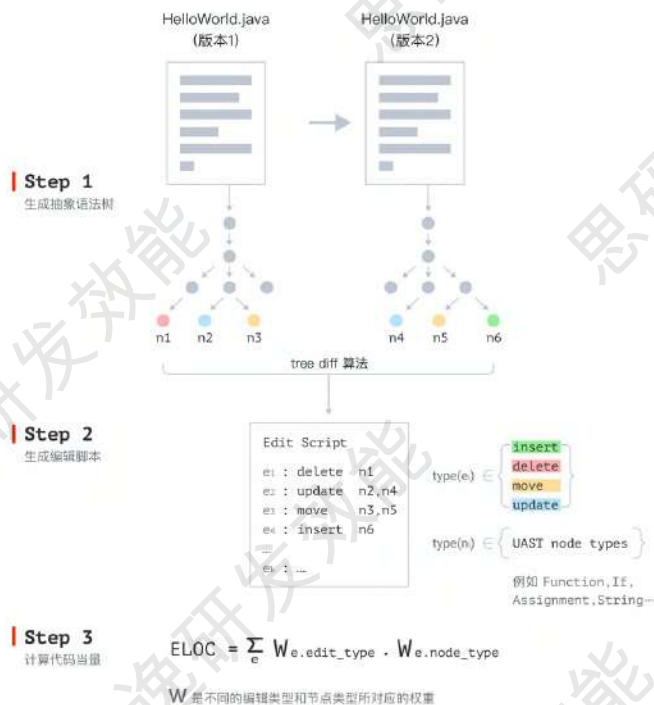
assignee_name	Feature Assigned (Finished/All)	Bug Assigned (Finished/All)	Other Issue Assigned (Finished/All)
warren830	18/19	10/12	5/6
abeizn	13/14	37/37	7/7
mappjzc	13/13	38/41	13/15
e2corporation	12/12	24/24	5/5
mindlesscloud	11/11	25/26	4/4
likyh	10/14	26/29	10/10
yuqiangabab	8/9	0/0	0/0



数据驱动研发效能的实践和案例 II

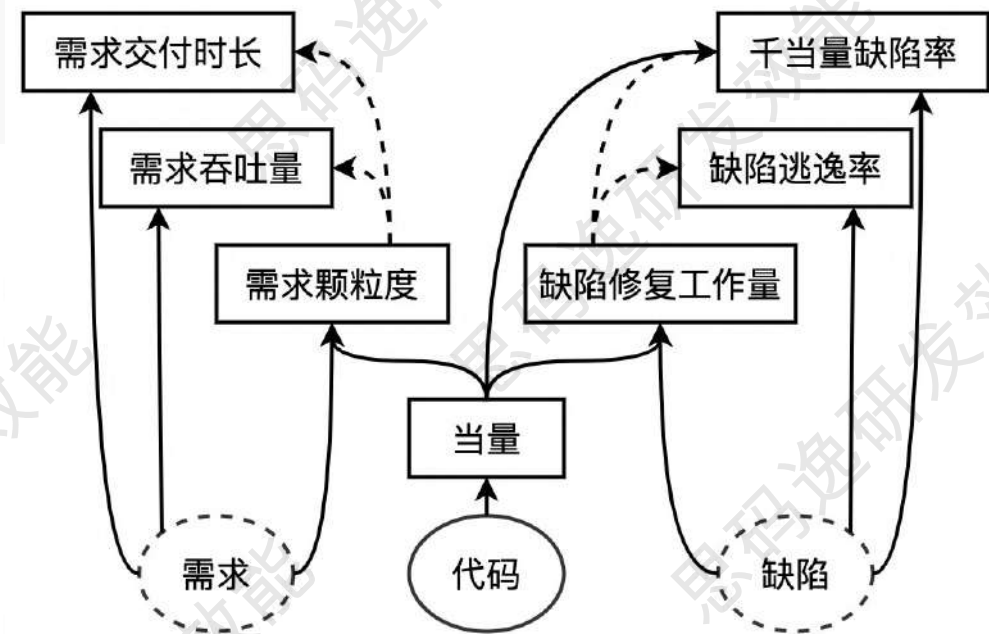
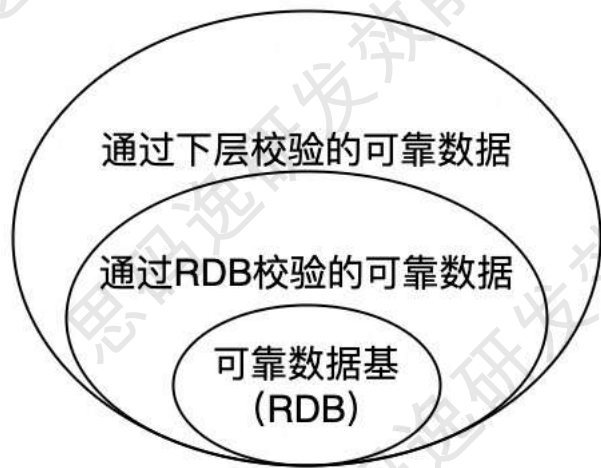
阶段	特征	陷阱	最佳实践
数据夯实	维护数据健康度，建立高质量数据源	缺乏可靠的数据基座，在指标上虚假繁荣	采用代码分析数据，并与其他领域数据交叉关联

如何计算代码当量(ELOC)



* 这里展示的是代码当量 (ELOC) 的基础原理，在实际运用中，开发当量的计算过程有更多的智能调节机制，例如对重复代码的调节。

分层数据校验



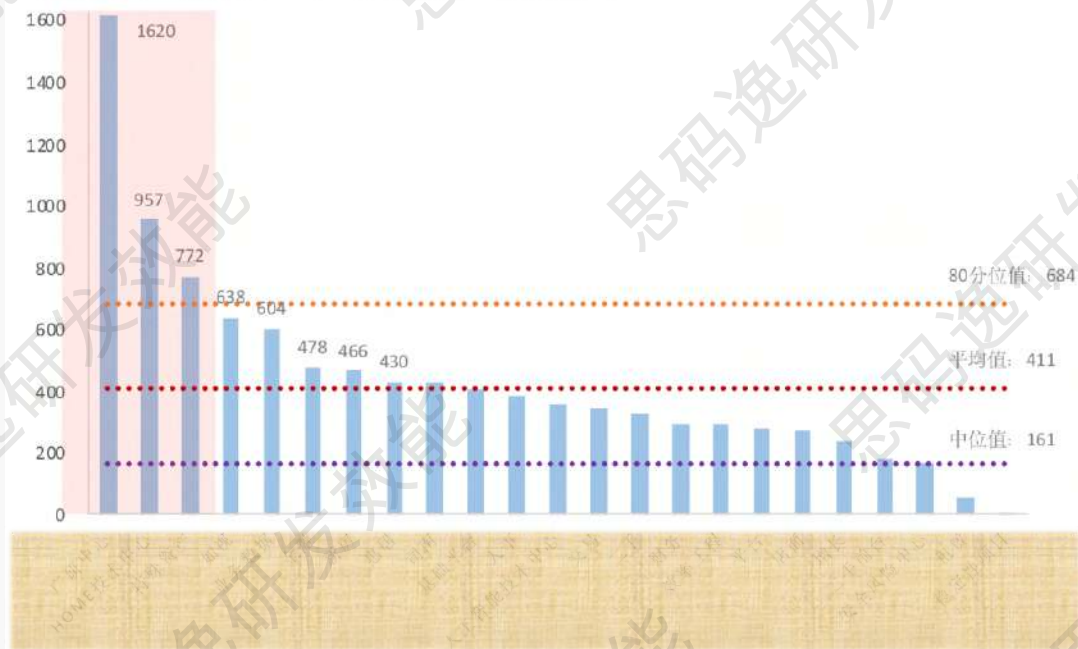
数据驱动研发效能的实践和案例 II

客户案例：
中国领先的新居住服务平台，美股上市公司



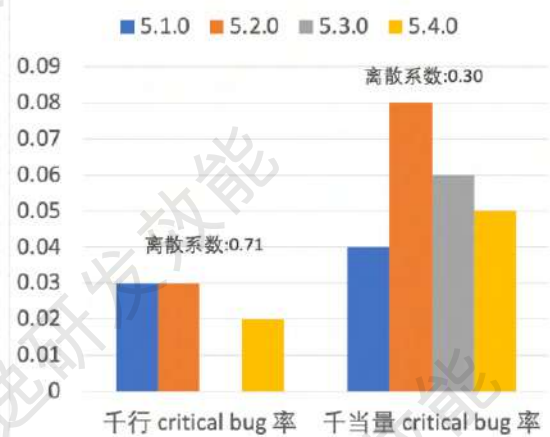
数据范围	数据跨度	样本数量
全部产品需求	2021.8-2022.4	34周

各业务产品需求 平均每需求当量排序

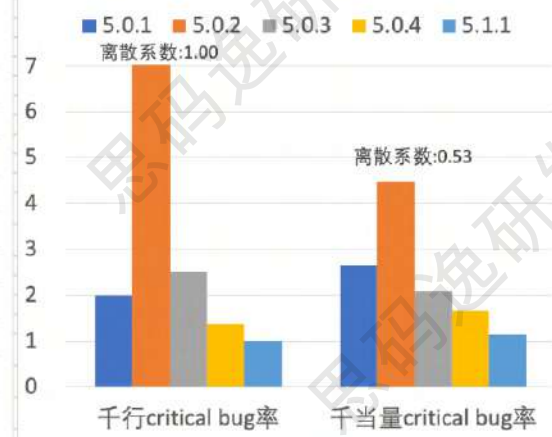


千当量 bug 率比千行 bug 率更稳定（离散系数低），更适用于评估版本质量。

minor 版本 critical bug 率



patch 版本 critical bug 率

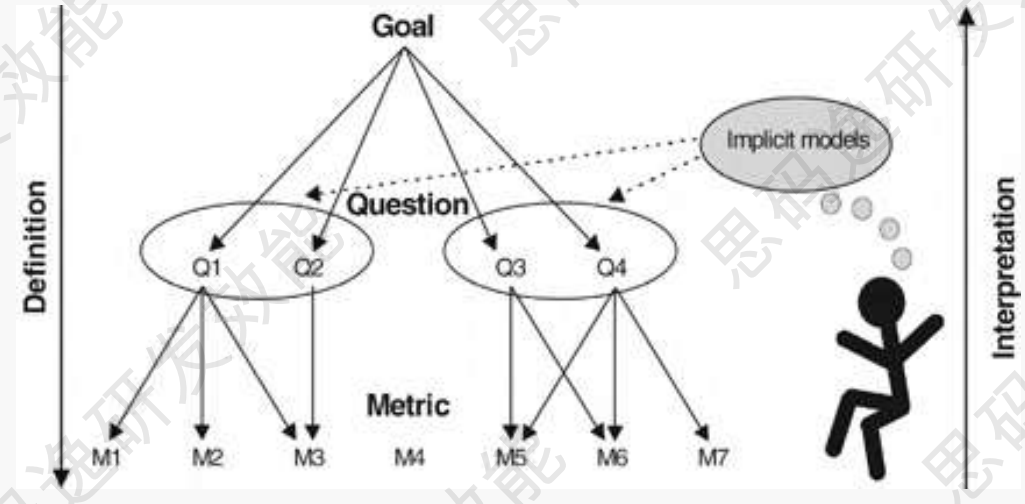


数据驱动研发效能的实践和案例 III

阶段	特征	陷阱	最佳实践
自主分析	利用BI或可视化工具，从数据中解读信息	围绕数据而非目标；缺乏适合研发模型的下钻，只能满足浅层信息需要	采用 GQM（目标-问题-指标）方法，通过合理下探推动 MARI（度量-分析-回顾-改进）循环

目标-问题-指标（GQM）是软件工程领域的一套系统方法，被称为研发效能度量方法的“事实标准”。

- 目标（概念层）
 - 举例：开发者改进代码合并请求处理的及时性
- 问题（操作层）
 - 举例：代码合并请求的处理速度如何？
- 指标（量化层）
 - 平均处理时长
 - 处理时长的离散系数
 - 超出处理时长上限的比例
 -

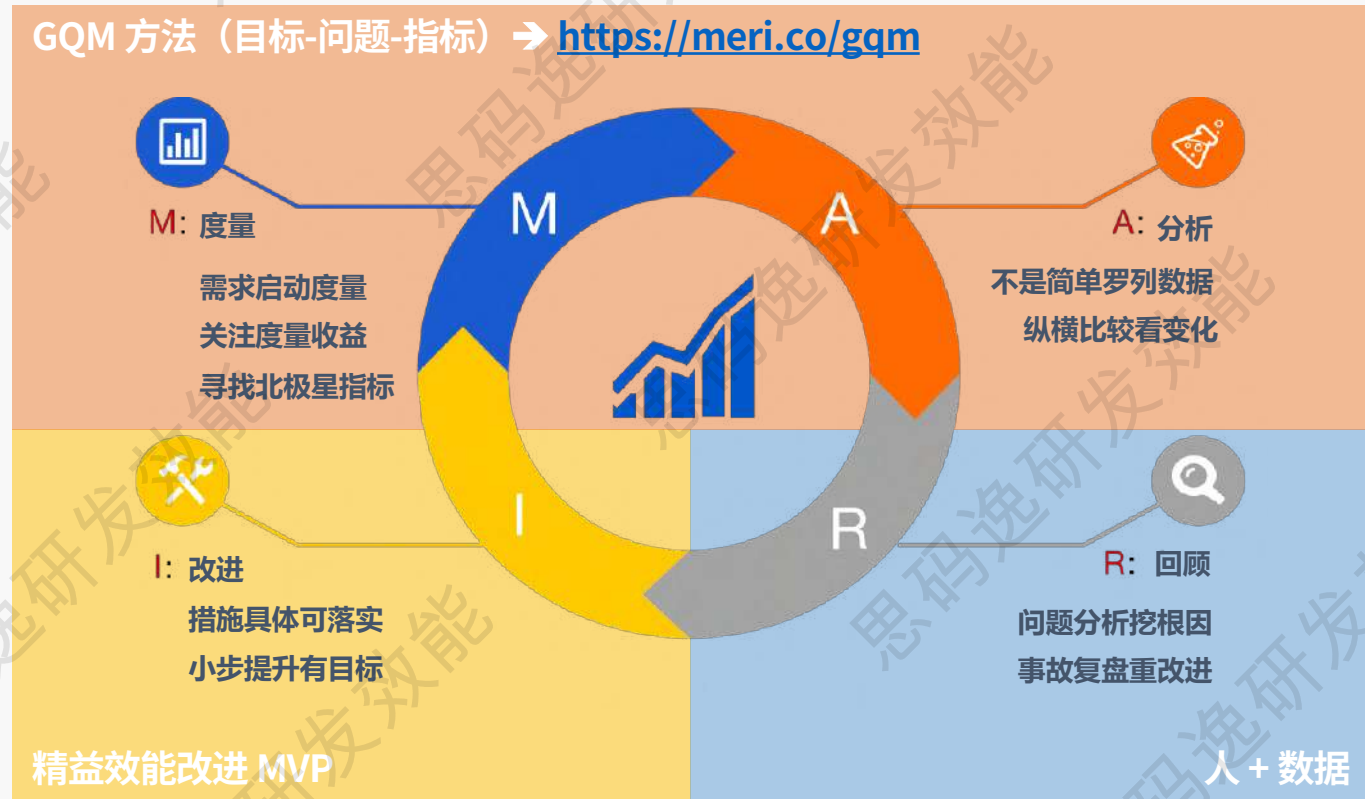


更多请参考《GQM 从入门到精通》 <https://meri.co/gqm>

数据驱动研发效能的实践和案例 IV

MARI 方法论的目的是建立从效能度量到改进的闭环。

GQM 方法 (目标-问题-指标) → <https://meri.co/gqm>



更多信息参见: <https://www.openmari.dev/>

问题: 问题本身难量化; 找不到好指标; 找不到好基线标准
目标: 沉淀专家经验, 建设专业、统一、标准、可工程化实现的“研发指标体系”

标准方法: 基于研发过程模型 + GQM方法设计

最佳实践: 字节跳动DevMind (2020→2021年) 研发指标流水线 标准指标 研发指标即代码

只有当人类“发明了发明的方法”之后, 人类社会才能更快地发展。
—查理·戴 DevOpsClub

问题: 分析每个场景严重定制化; 难以高频、持续做; 改进问题缺少跟踪
目标: 成为研发管理驾驶舱, 所有业务、团队在线化分析、决策、驱动改进

标准方法: 构建闭环, 持续改进

最佳实践: 字节跳动DevMind (2022年至今) 研发管理驾驶舱 (内置N份研发洞察报告)

度量是手段, 洞察是过程, 推动有效的改进或辅助管理者做出有价值的决策。—李目的 DevOpsClub

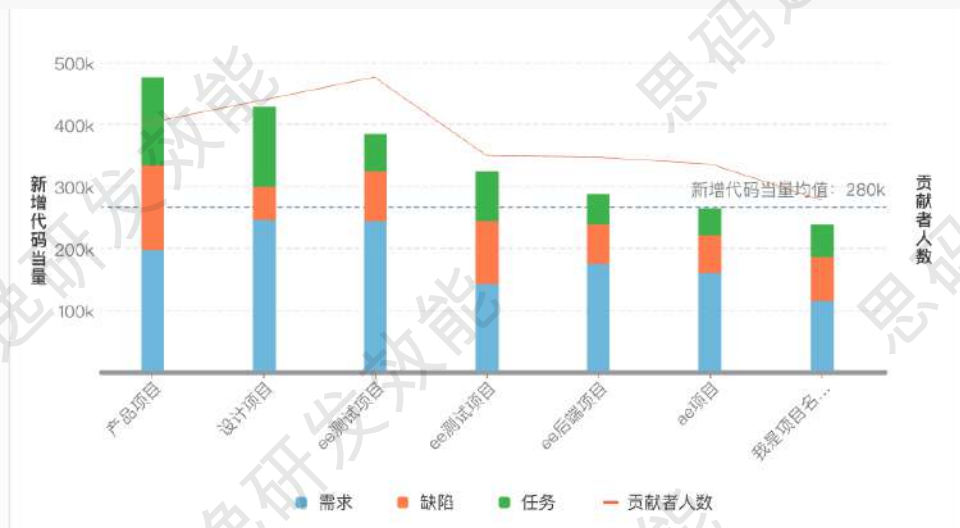
更多字节跳动的实践参见: <https://meri.co/8a323187>

数据驱动研发效能的实践和案例 IV

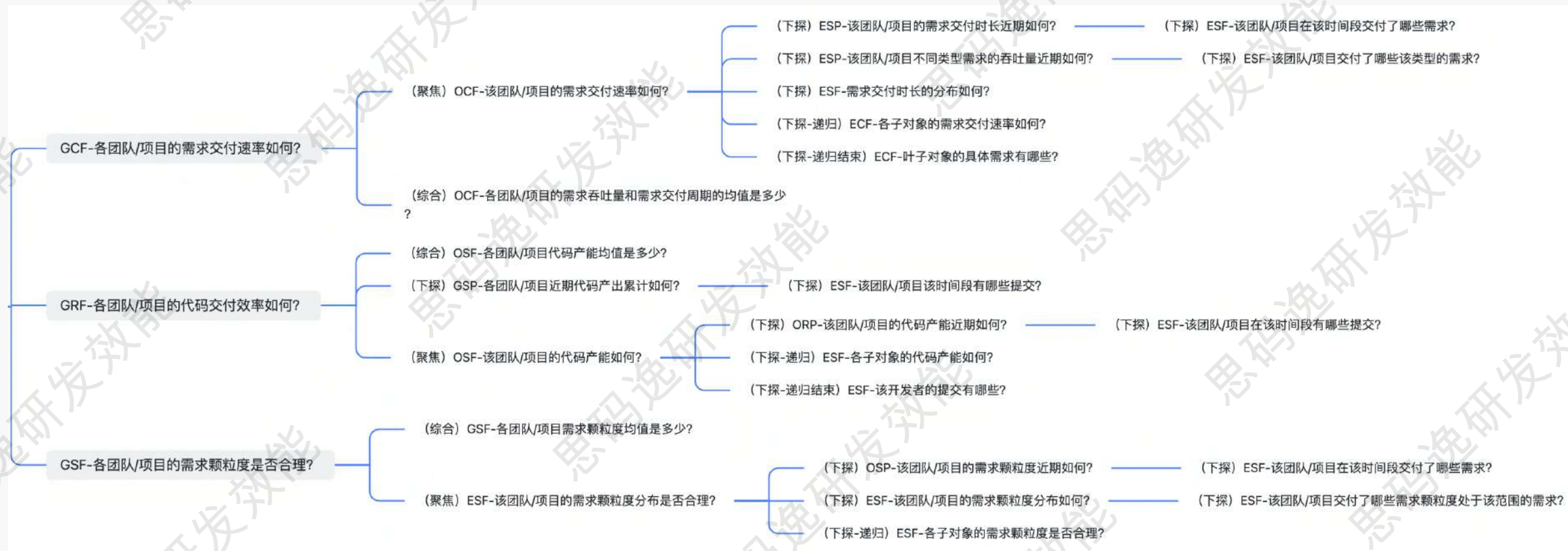
阶段	特征	陷阱	最佳实践
系统洞察	以目标为导向，基于健康的数据，合理分析和洞察	多样而有特异性的分析，强依赖分析师个人，无法在组织中规模化应用	参考规范化 GQM 模板，自主实现洞察和 MARI 循环；进一步复用和丰富模板，沉淀知识



GQM规范化方法参见：
<https://www.openmari.dev/docs/metrics-sys/gqm>



数据驱动研发效能的实践和案例 IV





提交免费试用

思码逸基于深度代码分析技术，为研发团队提供**一站式效能分析平台及解决方案**。

目前已为**腾讯、美团、滴滴、贝壳、自如、平安、泰康、戴尔EMC、商汤、蔚来汽车**等众多行业标杆客户提供服务。



研发效能交流群

在 *DevData Talks* 开放务实地聊聊研发效能



分享合集



研效交流群

💡 每月直播 干货满满

行业 KOL + 企业研发效能负责人 + 资深效能顾问
从效能建设路径到不同场景应用，分享前沿观点与先进实践

💡 开放探讨 共同成长

直播 Q&A + 交流社群
直播嘉宾面对面答疑解惑，交流群随时随地探讨效能话题

关于思码逸

作为 DevData Talks 的发起方，思码逸专注为企业研发团队提供效能数据分析，呈现效率、质量、人才发展等多视角数据洞察，辅助团队决策与工程改进。

思码逸已服务 腾讯、美团、滴滴出行、中国平安、泰康保险、戴尔 EMC 等众多行业标杆客户。