



DataEase

FIT2CLOUD飞致云 旗下品牌

2022 年 3 月版

新一代数据可视化分析平台 建设指南

HOW TO BUILD YOUR VISUAL ANALYTICS PLATFORM

DataEase 开源项目组 编著

从闭源到开源， 我所在的软件研发团队经历了什么？

2021年6月28日，DataEase 开源数据可视化分析平台正式对外发布。这是 FIT2CLOUD 飞致云在开源软件领域的新作，DataEase 的核心功能是通过支持丰富的数据源连接和提供拖拉拽的快捷方式帮助用户快速制作仪表盘，并且支持灵活方便的仪表盘分享机制。

作为一名软件开发行业的老兵，这是我第一次以开源的方式参与打造一款软件产品。

2014年 FIT2CLOUD 飞致云创立。作为创始团队的一员，我们看到了伴随着云计算应用普及而兴起的云基础设施管理需求，设计并实现了云管平台（Cloud Management Platform）软件，并且在后来的数年间不断地改进和升级这款产品。FIT2CLOUD 云管平台（后来被更名为 CloudExplorer 多云管理平台）是一款采用闭源方式开发的企业级软件。从2014年到2020年的六年多时间里，我们的研发团队都在用相对传统的方式去研发这款软件。目前这款软件仍然在升级和优化，在金融、制造、能源、交通、物流、房地产等行业拥有接近200家企业用户。

2021年的2月，DataEase 开源项目组成立。在闭源的世界里沉浸式体验了多年的我尝试去拥抱开源软件的世界。在开源领域有本著名的畅销书叫《大教堂与集市》，从2021年2月到今天，我们这些从闭源世界走向开源世界的软件开发者可能还没有体会到从“大教堂”到“集市”的剧烈反差，但是也的确收获了与以往截然不同的开发体验。

首先就是大量社区用户的使用反馈让我们在短时间内就拥有了庞大的产品需求池。事实上，由于云管平台软件是云计算时代的产物，我们在早期设计时同样也会大量收集用户的需求，从中快速提炼出通用的需求，并付诸实现。因此，对于用户现实需求的采集、识别、规划和实现，早已经是我们的日常工作的组成部分。但是 DataEase 项目组还是被开源社区的力量所震撼。

记得在 DataEase 开源数据可视化分析平台 v1.0 版本发布后的第三天，DataEase 开源项目就进入了 GitHub 趋势榜的主榜，并且在 Java 趋势榜中位列第一；项目发布一个月后的7月30日，DataEase 项目的 GitHub Star 数突破 1000 个；8月6日，GitHub Star 数突破 2000 个；8月30日，GitHub Star 数突破 3000 个。

在开源社区被广泛关注的直接结果便是我们收获了很多用户。DataEase 项目累计下载次数超过 10000 次是在 11月24日，距离这个项目被大众所知仅仅过去了五个月。快速积累的用户群体催生出大量的使用反馈和产品需求。我们每天都会收到大量的建议，我们从其中筛选出优先级最高的需求，将其规划到下一个版本的研发工作中。DataEase 项目的需求池从来不会干涸，每天从海量的需求中识别进入下一版本规划的部分，成为了一种幸福的烦恼。

第二，为了满足用户的期待，我们加快了软件发布的频率，而双方的高效互动让我们彼此找到了共同促进的节奏。对于闭源软件来说，三个月、半年，甚至一年发布一个版本都是很常见的。但是开源的世界不容懈怠，你必须对用户的反馈做出快速的回应，必须满足用户的期待。

和 FIT2CLOUD 飞致云旗下的 JumpServer、MeterSphere 等开源项目一样，DataEase 项目也保持着每月一个版本的发布频率。这可以说是我们与开源社区用户交互的一种“约定”。要打造一款成功的开源产品，就需要给用户以强大的信心。而保持快速、稳定的迭代周期，正是我们向社区用户传递信心的方式。有些功能即便当前的版本不满足、有缺陷，但是项目本身能够用自己的迭代节奏去优化和弥补，我们也从这一过程中坚定了自己做好产品的信念。这种不断强化的信念感是在闭源软件的世界里体会不到的。

第三，研发团队直面用户让我们的成就感和挫败感都来得更直接。在开发云管平台这款产品时，虽然它的应用场景是在云的环境中，但是软件的研发和交付方式却是相对传统的。在这款软件开发的早期，我们会直接与用户沟通需求，但是随着用户数量的增多和使用场景的千

PREFACE

序言

差万别，软件产品的研发和软件产品的交付会在两个组织内展开。其中，研发团队专注于通用功能的演进和优化，而交付团队则会深入了解用户的具体需要，并在通用软件落地企业环境的过程中附以定制化的功能实现。

在这种模式下，研发团队并不会和使用软件的用户进行深入的沟通，对于产品的改进更多的是从交付团队获取反馈。而当软件在一些客户的环境中获得了成功的使用，例如帮助用户节省了大量的计算资源，或者帮助用户大幅度地提升了 IT 资源交付的效率，研发团队所收获的喜悦可能远不如交付团队那么多。我们深知，我们所开发的通用软件能够成功地被用户所使用，有赖于交付团队与用户在需求面的大量对接，以及交付团队在功能面对于软件产品的完善和补充。

开源的世界则截然不同。DataEase 这款软件的亮点和短板我们第一时间就能够从用户侧获取反馈。这其中有很多的鼓励与期待，也不乏抱怨和吐槽。当用户提及 DataEase 与之前所使用的闭源产品相比较的优点时，我们感到非常开心。而用户在社区交流中所表达的负面情绪，我们也能欣然接受。社区用户对项目表达负面情绪，也可以理解为一种期待。通常情况下，在社区群中提出负面反馈的用户或多或少都会给我们的产品设计提供一些思路，我们也会尝试去理解和识别。

如果说闭源的世界是一面湖水，那么开源的世界就是一条奔涌的河流。

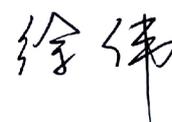
在开发云管平台软件时，版本发布的间隔时间会比较长，很长的一段时间内产品的功能、架构都是固定的。当这款软件经过足够长时间的迭代，其形态会进一步固化，而开源软件似乎总是处在变化之中。因为用户的需求是动态变化的，同时产品的 Bug 也需要及时修复。所以除了每个月发布的版本，我们也会根据需要发布小版本。可以说开源软件是被整个社区推动着进步，用户的需求便是其蓬勃成长的沃土。两者之间是“独自成长”和“开放成长”的区别，开源让软件产品拥有了更多的可能性。

前面谈的都是闭源与开源的话题，最后来谈谈我们为什么要用开源的方式来开发 DataEase 这款数据可视化工具。原因有两个：

首先，创立 DataEase 项目出于一家软件公司自身的需求。 FIT2CLOUD 飞致云旗下的 CloudExplorer 多云管理平台、JumpServer 开源堡垒机、MeterSphere 开源持续测试平台等软件在运营的过程中都会产生大量的数据，同时这些软件拥有大量的用户群体，用户普遍具有将这些数据进行可视化展示的实际需求；

其次，数据可视化市场广阔。作为一种跨领域、跨平台的产品，数据可视化软件具有强大的通用性，受众群体广泛，可以说是人人需要。如此庞大的市场发展了数十年，孕育了众多优秀的软件产品，通过开源模式进行研发和交付的软件却很少。我们尝试去用先进的设计思路、开源的软件分发方式，以灵活的软件架构去打造一款数据可视化平台，不断向先行者学习，根据用户需求持续改进，相信在红海中 DataEase 项目也能找到自己的市场空间。

FIT2CLOUD 飞致云在 2021 年提出了一个口号——“广阔天地，大有作为”。无论是开源，还是数据可视化，对于 DataEase 项目来说都是一番广阔天地，我们会努力在其中有所作为。



DataEase 产品总监 徐伟

CONTENTS

目录

1. 引言	1	4.3.1 简单易用的操作方式	17
1.1 编写白皮书的目的	1	4.3.2 开源开放的运营模式	17
1.2 如何组织、发布和维护白皮书	1	4.3.3 大数据下的秒级响应	19
1.3 本白皮书给企业带来的价值	1	4.3.4 安全分享	19
2. 数据可视化概述	2	5. DataEase 仪表盘案例集	20
2.1 数据可视化的作用	2	6. 参考资料	23
2.2 数据可视化的价值	3		
2.2.1 数据可视化对企业的价值	3		
2.2.2 数据可视化对不同人员的价值	4		
2.3 数据可视化的发展	4		
2.3.1 可视化简史	4		
2.3.2 现代可视化发展	5		
3. 新一代数据可视化分析平台的内涵	6		
3.1 简单易用	6		
3.2 开源开放	7		
3.2.1 开源模式更好地推动软件产业发展	7		
3.2.2 开源生态发展现状	8		
3.3 高性能数据引擎	10		
3.4 快速分享	11		
4. 基于 DataEase 构建新一代数据可视化分析平台	12		
4.1 DataEase 概述	12		
4.1.1 DataEase 架构设计	12		
4.1.2 DataEase 核心概念	12		
4.2 DataEase 的功能列表	15		
4.3 DataEase 的核心优势	17		

1. 引言

1.1 编写本白皮书的目的

数据的采集、提取和理解是人类感知和认识世界的基本途径之一。相对于数字、文本等非图像化信息，人类对图像化信息有着更好的感知能力。基于这个原因，数据可视化逐渐成为了人们更深入分析理解数据、获取知识并转化为智慧的一种主要途径。数据可视化分析的发展史已逾数百年。自个人计算机问世以来，可视化基础平台和架构、常用可视化分析软件就在不断被研究开发，研究的领域包括医学可视化、科学可视化、信息可视化等。

目前国内很多企业的数据可视化和报表制作流程，是以业务部门提出需求后 IT 部门再进行开发的方式为主，这就需要多部门间的配合，存在着效率低下的问题。也有一些企业选择了内部研发，或者从外部引进可视化分析项目的方式。但事与愿违的是，很多可视化分析系统并未真正在用户手中使用或者传播开来。究其原因，这些可视化分析系统并未给用户带来足够的便捷性，甚至还比不上用户原本使用的文档类工具。

DataEase 开源项目 (www.dataease.io) 在成立之初就致力于构建一个人人可用的开源数据可视化分析平台。正是带着这样的目标，DataEase 开源项目选择了不断地听取社区声音、持续地进行改进与迭代的路线。在这一过程中，DataEase 开源项目逐步有所收获，并快速成长。希望本白皮书可以为正在或即将建设数据可视化分析平台的企业提供一些实际的帮助。同时，DataEase 开源项目也希望能够通过白皮书的方式和大家进行更多的沟通和交流，共同探讨应该如何更好地建设新一代的数据可视化分析平台。

1.2 如何组织、发布和维护本白皮书

本白皮书由 DataEase 开源项目组联合编写而成。编写团队从数据可视化分析领域技术发展的角度出发，站在用户的视角，积极吸取行业内成熟产品的经验和专家的建议，在此基础上完成了本白皮书的编写工作。

本白皮书采用线上渠道（网站、微信公众号等）为主的分发模式。在第一版线上发行后，编写组将通过线上渠道广泛收集用户反馈并定期更新白皮书内容。更新版本的白皮书仍将通过线上渠道再次对外开放下载。欢迎广大用户及各位业界同仁积极提交修正和改进意见，协助编写组对本白皮书进行持续完善，使之更加准确、全面和深入。

1.3 本白皮书给企业带来的价值

本白皮书面向的人群为广大需要使用或建设数据可视化分析平台的团队，包括但不限于业务部门、运营部门、产品部门、IT 部门等。编写组希望通过本白皮书为企业建设数据可视化分析平台提供方向上的指引，并希望更多的数据可视化分析工具能够被广泛使用起来，真正地走进用户群体，最大化地发挥软件的价值。

由于数据可视化分析平台本身是一项庞大的课题，本白皮书未能详尽地涉及数据可视化分析平台的方方面面。编写组的主要目的是起到抛砖引玉的作用，通过制作与分享本白皮书，引发大家对数据可视化分析的更多思考与讨论，共同为新一代数据可视化分析平台的建设发掘更多的思路。

2. 数据可视化概述

可视化是指通过创建图形、图像或动画的方式来交流沟通信息的技术和方法，历史上的洞穴壁画、埃及象形文字等都属于可视化的范围。如今可视化的应用领域不断扩大，覆盖科学教育、工程、互动多媒体、医学等领域。随着计算机科学的发展，可视化在计算机领域特指利用计算机图形学和图像处理技术，将数据转换成图形或图像并在屏幕上显示出来，再进行交互处理的理论、方法和技术。

2.1 数据可视化的作用

▶ 可视化有助于人眼接收信息

人眼是一个高带宽的巨量视觉信号输入并行处理器，具有很强的模式识别能力，对可视符号的感知速度比对数字或文本快多个数量级，且大量视觉信息的处理发生在潜意识阶段，被称为先意识处理。例如大脑会倾向于把一个复杂的物体解析成较为简单的物象来理解，从而降低认知负荷。著名的视觉设计格式塔原则便是利用了人类在视觉感知上的各种倾向性。总的来说，基于人眼的特殊性，我们可以认为可视化的含义是“生成符合人类感知”的图像，从而使人类更有效地接收信息、形成结论并传播知识。

▶ 可视化有助于探索性分析

探索性数据分析是利用各种技术手段探索数据内部结构和规律的一种数据分析方法和理念。其目的通常为理解数据并找出其中值得关注或有价值的信息，即将“数据”转换为“知识”的过程。在探索性数据分析过程中，用户是行为的主体，通过将数据可视化后使得视觉感知器官更好地获取信息、编码并形成认知，在数据可视化交互分析过程中获取解决问题的方法，或找到相关问题的原因。探索性分析是数据分析师工作的重点之一，通过将数据转换为图像，化繁为简，激发深度思维，再进一步使用搜索、过滤、联动、上卷、下钻等操作辅助探索数据中有价值的信息。

▶ 可视化有助于解释性分析

解释性分析是指主体将已经发现的精华信息抽象出来，再通过展示、描述、沟通、演讲等方式向其他对象传递的过程。《数据之美》一书将可视化定义为一种媒介，你可以通过可视化展示趋势、对比、离散值等图像来帮助他人“看到数据”，帮助接受解释性分析的人注意到他们从来没有想到会看到的内容。另外，古老的象形文字、丰富的壁画、历史交通运输图、行军图、人口统计图等历史图像，可以为人们解释一些过去的人、事和文化，成为我们了解历史、了解祖先的重要途径之一。



图1 婚礼照片数量的可视化

2.2 数据可视化的价值

数据可视化技术也常常作为 BI (Business Intelligence, 商业智能) 工具的一部分被探讨。BI 即由数据仓库 (或数据集市)、查询报表、数据分析、数据挖掘、数据备份和恢复等部分组成, 以帮助企业决策为目的的技术及其应用。事实上, 目前国际前沿的 BI 产品已具备一些机器学习、智能预测、AI 问答方向的功能, 那么为什么我们仍然还要讨论数据可视化的价值呢?

根据 Gartner 在《市场趋势: 中国企业买家越来越喜欢通过本地供应商进行虚拟化, 数据和分析》报告中的描述, “对传统报表的高度关注意味着专注于新兴能力的非中国供应商与中国企业的需求不符。大多数中国企业仍在寻找一款 BI 工具来执行描述性分析, 而不是诊断性或预测性分析。”

而 IDC 在《2021 上半年中国商业智能软件市场数据跟踪报告》中也提到, “在商业智能软件的细分子市场中, 报表分析软件市场仍是目前中国市场最主要的需求。2021 年上半年市场份额占比为 84.1%, 市场规模 2.7 亿美元。”由此可以看出, 基于现阶段中国企业的需求现状, 可视化分析方向的能力与基础设施建设才是我们当下更应该关注和聚焦的。

2.2.1 数据可视化对企业的价值



图2 数据可视化对企业的价值

► 解决数据孤岛，把握市场发展趋势

随着信息化、数字化的普及与建设, 各领域各企业也随之不断完善数字化产品和服务。在新冠疫情的影响下, 企业的数字化转型意愿不断加强, 数字化转型需求持续拓宽。然而在数字化转型的过程中, 时常会出现数据壁垒过高、不易流通、联动困难、易形成数据孤岛等问题。业务部门仍然依赖于 Excel 做数据的流转和处理, 而 IT 部门的工作重心则需要集中在为业务人员提供查数、取数, 以及开发报表展示数据等服务方面。

数据孤岛的存在显然不利于企业展开经营分析, 同时也加大了挖掘数字价值的难度。企业需要统一的工具进行存数和取数, 从而打破数据壁垒, 释放 IT 人员的生产力, 提升部门效率, 并且能够更好地在数字化转型与发展的过程中最大化地利用数据的价值, 进而找准业务方向, 把握市场机会。

► 加强用户行为管理，强化用户协作

数据可视化的作用在于格物致知, 需要达到真善美的均衡, 即达到数据的真实性、表达意向的倾向性、表达形式与内容的和谐统一性

的均衡。通过文件导入导出、层层传递、上报数据的方式却与这一理念相悖，并且无法很好地保证数据在各人员协调过程中保持真实统一性与安全性。

企业在解决数据孤岛问题的同时，还应该思考如何针对员工的数据处理、流转、可视化、分享等过程进行管理。首先，我们需要建立完善的数据接入及使用权限体系，合理下放数据使用权限至部门与员工；其次，降低员工对于数据分析或数据报表结果分享操作的难度。使用合适的可视化分析工具取代文件层层上报、汇总的方式，有利于用户协作与数字化建设的交叉发展。

► 建立指标体系，实时分析规避风险

企业可以通过分析 KPI、业务运营情况、业务流程等数据，梳理各项指标，从不同维度评估业务，建设符合自身发展的指标体系。企业通过可视化指标体系，并且通过监控指标仪表盘对业务流程整体与局部的情况进行实时监测，在监测过程中提前评估业务中可改进的地方，从而优化部署，规避可能的风险。

2.2.2 数据可视化对不同人员的价值

前端业务人员：能够快速、高效地提供体验良好的数据展现手段，通过分析数据资产的质量，助力企业做出更加准确的业务策略；

数据管理人员：提供多种数据使用模式，提供更为丰富、安全的数据管理手段，有助于企业内部进行更为广泛的数据整合与分析，并由此创造数据价值；

IT 建设人员：能够规范数据处理的过程，保证数据质量，提升 IT 系统建设效率，快速支持业务部门的运营发展需要，有效激发 IT 建设的创新力，体现 IT 信息的价值；

企业管理者：能够合理评估、规范和洞察企业信息，洞悉企业发展趋势，在有效管理 IT 投资和降本增效的同时，挖掘和发挥数据资产的价值，辅助企业制定发展决策。

2.3 数据可视化的发展

数据可视化早已体现在现代社会的方方面面，街边的广告、各类报告上的分析图表、工作汇报总结等内容都是数据可视化的体现。身处现代社会的我们早已习惯了身边的种种可视化表达，那么可视化理念是从何时开始在人类的生产生活中应用与发展的呢？

2.3.1 可视化简史

早在 17 世纪之前，图表在人类社会中就on已经开始萌芽，人们使用手工方式制作一些可视化物品来展示信息。在后期的发展过程中，可视化逐步开始应用于物理测量、图形符号、人口统计、政府规划、统计学等领域。

18 世纪，William Playfair 发明了折线图、柱状图、饼状图等在我们现代社会中时常用到的重要基础图形。19 世纪上半叶，统计图形、

概念图等图形发展迅猛，人们掌握了整套的统计数据可视化工具。同时，关于社会、地理、医学和经济的统计数据也越来越多。随着统计图表的持续发展，人类进行可视化思考的方式也不断推陈出新。

20 世纪，个人计算机的普及使得人们逐渐开始采用计算机编程的方式实现可视化。20 世纪 70 年代以后，桌面操作系统、计算机图形学、图形显示设备、人机交互等技术的发展，激发了人们通过编程实现交互式可视化的热情。

资料来源：《数据可视化》



图3 可视化简史

2.3.2 现代可视化发展

进入 21 世纪后，原有的可视化技术已难以应对海量、高维、多源和动态的数据分析挑战。大数据、数据分析行业迎来了高速发展的时期，国家与企业更加重视数据价值，强调数字化建设的重要性。人类需要综合可视化、图形学、数据挖掘以及新的理论模型、用户交互手段等技术，辅助现代社会用户应对当下的挑战。可视分析学应运而生。

可视分析学综合了图形学、数据挖掘和人机交互等技术，以可视交互界面为通道，将人的感知和认知能力以可视的方式融入数据处理过程，形成人脑智能和机器智能的优势互补，建立螺旋式信息交流与知识提炼途径，完成有效的分析推理和决策。



图4 可视分析学涉及的学科

从 2004 年起，研究界和工业界都在沿着面向实际数据库、基于可视化的分析推理与决策、解决实际问题等方向发展。与此同时，软件行业也受到推动，可视化的基础平台和架构、常用的可视化软件和系统迅速发展，并且按照可视化对象划分为医学可视化、科学可视化、信息可视化和可视分析四个类别，在各行各业中帮助人们解决实际问题。

3. 新一代数据可视化分析平台的内涵

随着科技发展、软件产业的成熟，数据可视化分析类软件已经稳步发展了一定的周期，并在国际与国内都形成了庞大的市场。然而与之不匹配的是，数据可视化分析基础设施建设并未在中国流行与普及，或是相关系统建设落地后却没有带来应有的使用率。

究其原因，是需求窗口小吗？从上文对数据可视化价值与意义的分析中可知，可视化的流量相当巨大。但很多用户反馈，使用了数据可视化分析软件后，并未让自己的工作变得更加智能，因为其繁琐的部署和升级操作、华而不实的功能、较高的上手门槛等因素，让用户又转而回到了 Excel 的怀抱。那么基于中国当下的市场、用户和企业应用现状，我们该如何构建新一代的数据可视化分析平台呢？

3.1 简单易用

在做可视化分析平台的选型时，我们会谈论些什么呢？站在一名高级数据分析师的角度，希望该产品具备丰富、前沿的功能，能够有效地引领数据分析工作；站在一名公司采购人员的角度，希望该产品有更高的性价比；站在一名运维人员的角度，则希望该产品更具稳定性。当然，以上要素都是一家企业，甚至某个单一用户在横向测评相关软件时会去考量的因素。但在众多影响要素中，哪一类是权重最大、排在人们第一序位的呢？知名研究机构 Gartner 在《Survey Analysis: Key Selection Criteria for Business Intelligence and Analytics Platforms》报告中曾发布过一份调研报告。报告显示，对于商业智能分析软件工具，易用性是企业获取商业价值排在第一位的影响要素。

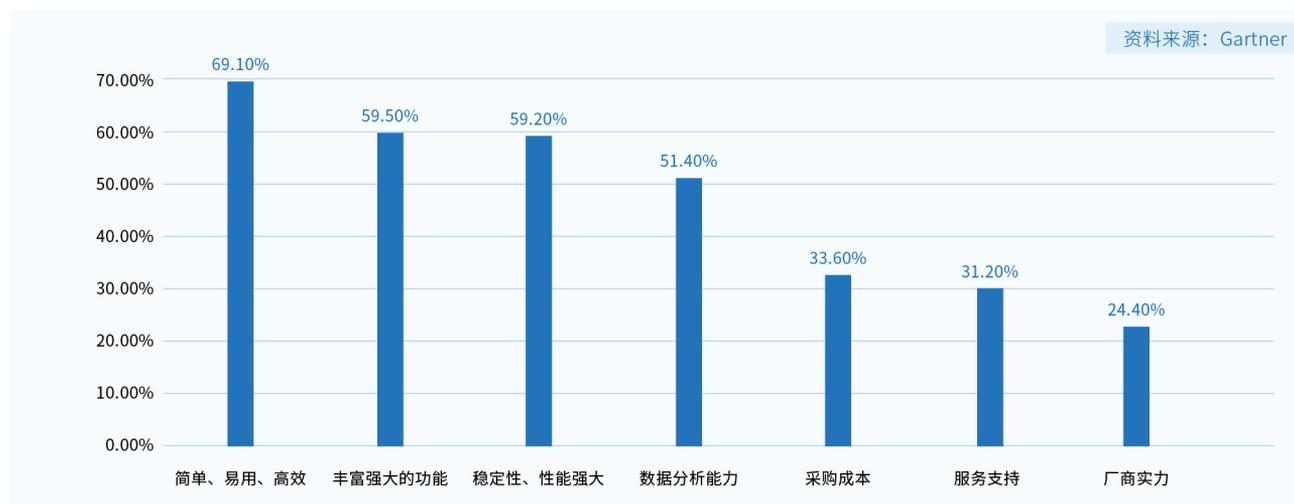


图5 企业选择商业智能分析软件工具的影响要素

► 获取方便安装简易

获取安装是用户使用一个工具的第一步，而这也是容易被软件的开发者所忽略的一步。方便快捷地获取各版本对应的安装包、简易的安装方式、通用的升级步骤，可以在使用者心中建立初步接触的好印象。一方面，使用者可以快速进入到下一步的实际功能操作中；另一方面，对于系统后期的维护与升级也不会有太大负担；

▶ 功能清晰易上手

数据可视化分析的重点在于提出问题和解决问题，而非学习软件。新一代数据可视化分析平台在设计时，应当始终以用户为中心，重视用户操作体验，简化系统功能的流程。使用者无需具备专业技术功底，或是依赖于其他 IT 人员，只需要专注于自己要解决的问题即可，避免花费过多精力在软件的适应与学习方面；

▶ 丰富的学习参考资料

优秀的可视化分析平台还应配套丰富的线上文档、教学视频等，在用户需要时提供更加全面的帮助，例如使用技巧、软件版本更新详情、主题仪表盘分享、常见模板分享、数据分析案例分享等。同时基于该软件具有类型广泛使用对象的特性，数据分析平台还可以搭建统一的用户交流渠道，加强使用者之间、使用者与平台运营者之间的沟通交流、经验互鉴。

3.2 开源开放

近年来开源产业发展迅速，在基础设施服务、数字化产业建设、推动产业协作、技术创新上均发挥了可观的效益。那么，开源开放的运作模式会带来哪些好处？目前开源产业的发展趋势又如何呢？

3.2.1 开源模式更好地推动软件产业发展

红帽在题为《2021 年企业开源现状》的研究报告中提到，如今 90% 的 IT 主管都在使用企业级开源软件。其中被首选使用的企业级开源软件的前三位类型分别为：基础设施现代化软件、应用程序开发软件和数字化转型软件。在受访者中，位列第三位的数字化转型软件占比为 54%，这是因为很多新型应用是数字化的重要组成部分，数据可视化分析平台便是其中的重要一员。

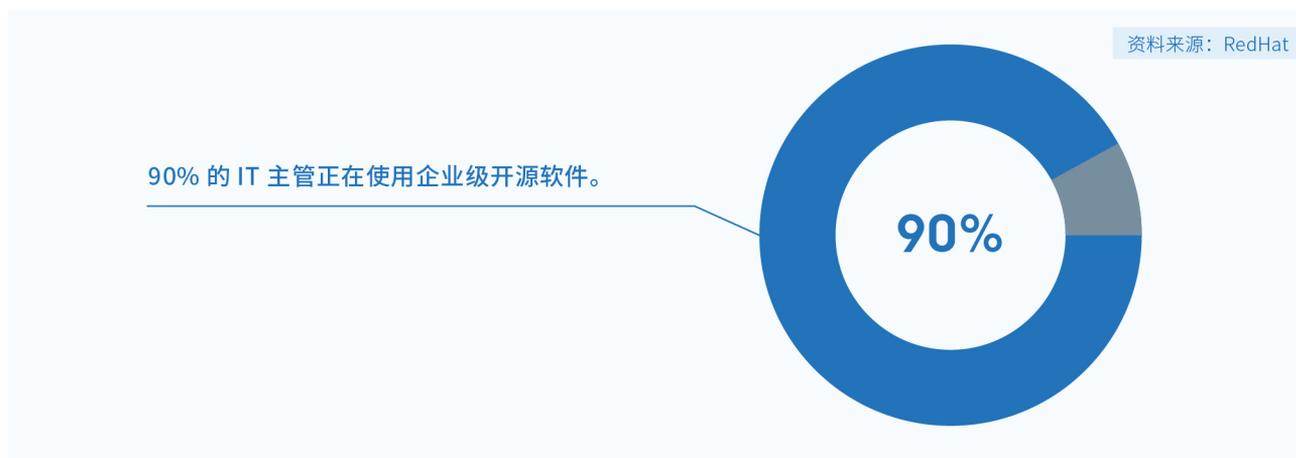


图6 90% 的 IT 主管使用企业级开源软件

► 开源模式推动产业发展

企业及个人依托开源生态进行交流与技术协作，打破了软件开发者与使用者之间的壁垒，通过及时的反馈与交流、快速分发、共同协作加速了软件的创新与发展。纵观数据可视化分析领域的软件发展史，其中有不少软件以开源的形式进行开发运营，包括开源医学影像分析与可视化开发包 VTK 和 ITK、开源跨平台的医学图像分析与可视化软件 Slicer、面向科学数据和工程数据的开放可视化环境软件 OpenDX 等。这些软件产品通过开源方式进行共享，成为推动可视化分析产业发展的重要动力；

► 开源运营可以听到和采纳更多的声音

以开源模式开发运营的软件，软件开发者可以通过代码托管平台、社区、社群等途径快速建立与使用者之间的联系。用户可以即时将自己使用的感受、建议意见反馈给开发者。与此同时，使用者也可以变成参与者，将自己的经验转换为功能代码提交给开源项目；

► 开源可以更好地促进产品快速迭代与验证

基于开源的运营模式，产品可以在迭代过程中快速聚集庞大的使用者或参与者。收集社区用户使用后的反馈意见、需求后，开源项目的主导者可以根据需求所对应的群体规模判断其产品研发的优先级，并按时升级迭代。每个版本发布后，开源用户也可以第一时间获取和升级使用验证，从而形成正向的反馈循环。此外，数据可视化工具对于行业或职位没有特殊的倾向性，涉及到各行各业，人员基数庞大。大规模的软件使用群体在开源社区的环境下可以更好地相互交流、汇总信息，并反哺产品的开发演进；

► 开源让软件更安全

开源软件对外开放源代码及免费安装包，这一形式让软件本身处在很多隐形的同行评审与用户评审中，时时刻刻受到广大开发者与使用者的监督，让缺陷无处遁形。对于发现的软件问题也能够第一时间处理发布，即时通知社区用户进行升级，将软件缺陷所能引发的系统风险降至最低。

3.2.2 开源生态发展现状

► 国际开源生态发展现状

据 GitHub 官方 2021 年度报告显示，目前 GitHub 在全球已经拥有超过 7300 万的开发者用户。在新冠疫情仍然持续发酵的 2021 年，全球开发者更加注重智能和可持续化的编码方式、文档创建、社区支持等开源生态功能。

同时，科技巨头企业也同样重视对外开源生态建设，并为其投入了大量人力。根据 Open Source Contributor Index 公布的 2020 全球开源厂商 GitHub 开源贡献排名，谷歌和微软两大互联网巨头排在榜单的前二位，二者参与开源的活跃贡献者人数均超过了 5000 人，参与的开源社区超过 10000 个。微软、谷歌、IBM、Oracle、Facebook 五家科技巨头的企业级开源项目数超过 20000 个。在软件使用占比方面，开源软件也毫不逊色。DB-Engines 的数据显示，截至 2021 年底，全球 383 款数据库中，开源数据库占据了 51.7%，高于 8 年前的 35.5%；排名前十的数据库中，开源数据库占据六席。

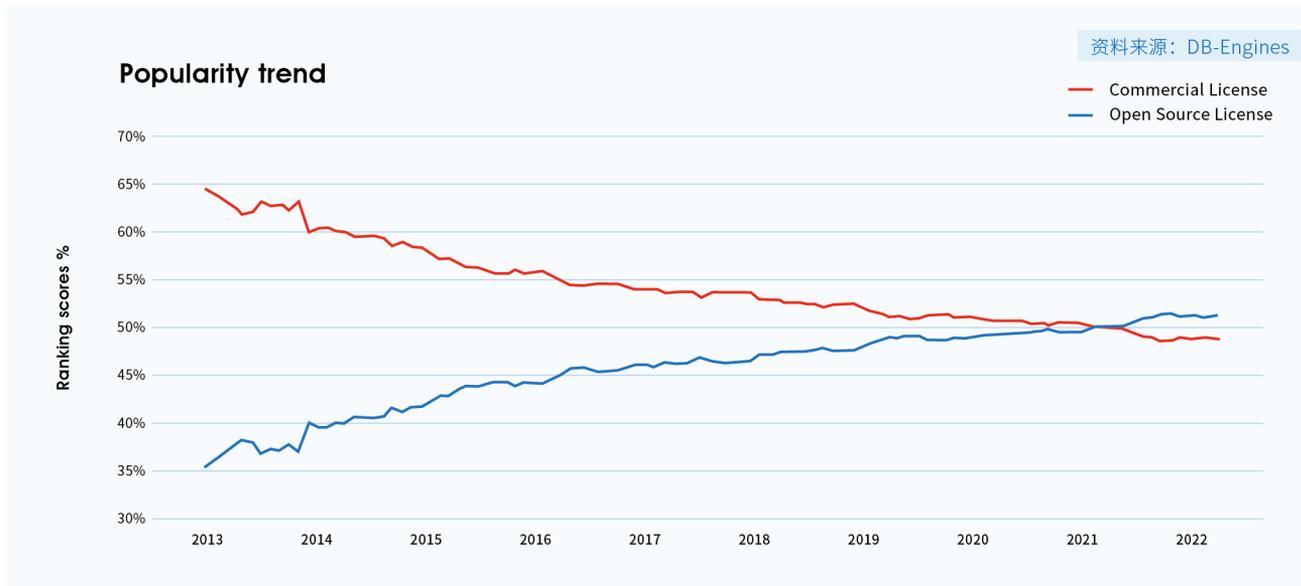


图7 全球数据库流行趋势

► 中国开源生态发展现状

在国际开源生态稳步增长的形势下，中国市场的开源脚步也在加快。近年来，中国的开源软件贡献者数量快速增长，在全球贡献者占比中不断攀升。GitHub 发布的 2021 年度数据报告显示，在全球已有的 GitHub 开发者中，来自中国的开发者数量以 755 万位列排行榜第二，仅次于美国。而在 Gitee 2021 年注册的开发者中，来自中国的开发者数量超过了 180 万，累计超过 800 万。无论是在全球市场还是中国市场，中国的开源软件开发者都呈现出加速增长的趋势，并且逐渐成为中国开源产业的基础力量。



图8 2021 年 GitHub 全球用户分布

► 国家政策方针指导

2021年3月颁布的“十四五”规划纲要在第十五章《打造数字经济新优势》中特别提出，“支持数字技术开源社区等创新联合体发展，完善开源知识产权和法律体系，鼓励企业开放软件源代码、硬件设计和应用服务。”这也是我国政府首次将开源纳入顶层设计之中。除此之外，中国人民银行、中央网信办等五部门在2021年10月发布了《关于规范金融业开源技术应用与发展的意见》，从多维度阐述了金融机构应当如何更好地利用开源技术，参与开源生态建设与产业发展。

3.3 高性能数据引擎

自2014年国家战略首次引入大数据技术开始，“数字”、“大数据”等关键词多次被写入国家战略文件，要求企业重视数据价值、促进数据应用、明确数据要素地位、强调释放数据要素潜能。在企业不断践行国家政策，进行自身数字化转型的过程中，数字经济规模不断攀高。

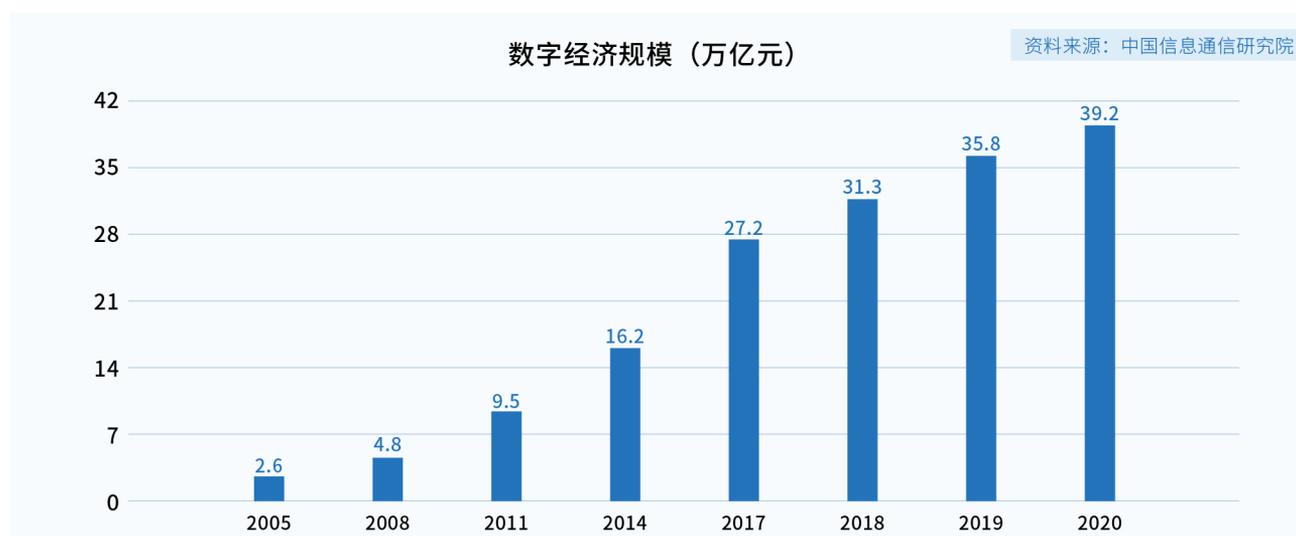


图9 我国数字经济规模发展趋势

这意味着即便数据的类型日益繁多，但是只是沉淀在数据库里是没有意义的。只有经过采集、抽取、展现成各种报表，给决策层提供参考才能为企业带来业务上的价值。然而这些多维、大量的数据在存储、分析、可视化上都给企业带来一定的挑战。因此，为更好地支撑现今数据规模与多元化的分析场景，新一代数据可视化分析平台的数据引擎应当具备以下特征：

► 足够高效

数据引擎需匹配当下数据发展的规模，在存储、执行、查询方面具备良好的效能；在存储上可以支持TB、PB级别，实时分析过程中达到秒级或亚秒级的效率；在数据分析查询的能力上为企业添砖加瓦；

▶ 简单易扩展

数据引擎的本质功能是为了提升企业在数据分析和决策方面的效率，而如果数据引擎本身在使用和运维方面有着极高的成本，那么即使这款数据引擎的性能再优异，依然无法给企业带来整体上的效能提升；

▶ 满足多种数据分析需求

新一代数据可视化分析平台应当具备固定历史报表、实时数据分析、交互式数据分析和探索式数据分析等支持多样化场景的功能，这就要求数据引擎在各类数据场景下都能具备良好的性能，让企业的数据分析工作更加简单高效。

3.4 快速分享

在数据分析可视化场景中，仪表板制作只是其中的一个步骤。将制作好的仪表板进行分享，用于工作汇报、解释性分析等工作，并分享给同事、领导，才能更好地发挥数据可视化的价值。因此，新一代数据可视化分析平台应该具备方便快速的分享操作方式，同时需兼容 Web 端、移动端等，满足不同人群对于数据分析结果的观看需求。便捷的分享过程还需注意兼备数据安全，保障敏感内容的私密性。

4. 基于 DataEase 构建新一代数据可视化分析平台

4.1 DataEase 概述

DataEase 是一款人人可用的开源数据可视化分析工具，帮助用户快速分析数据并洞察业务趋势，从而实现业务的改进与优化。DataEase 支持丰富的数据源连接，能够通过拖拉拽的方式快速制作图表，并可以方便地与他人分享。

4.1.1 DataEase 架构设计

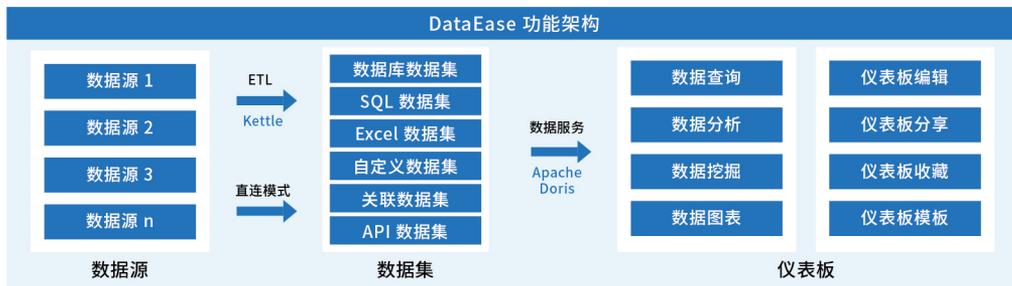


图10 DataEase 架构图

DataEase 系统的核心功能分为四层，将接入数据源到完成仪表盘的过程，合理归类为三个独立模块。这样一来，操作流程清晰易理解，各部分元素可多场景复用。除了核心功能外，系统还具备企业级别的权限管理、用户管理、租户管理、角色管理等功能，具备完善的 REST API，支持单点登录、嵌入式集成等。

4.1.2 DataEase 核心概念

DataEase 工作原理如图 11 所示。接下来，我们通过进一步地理解 DataEase 的核心概念来深入地认识 DataEase 的工作流程。



▶ 数据源

数据源即数据来源，管理员可在此配置数据库相关连接信息，校验数据库的连接性。完成数据源的配置后，可在后续数据集操作中使用，作为搭建用户与数据之间桥梁的第一步。

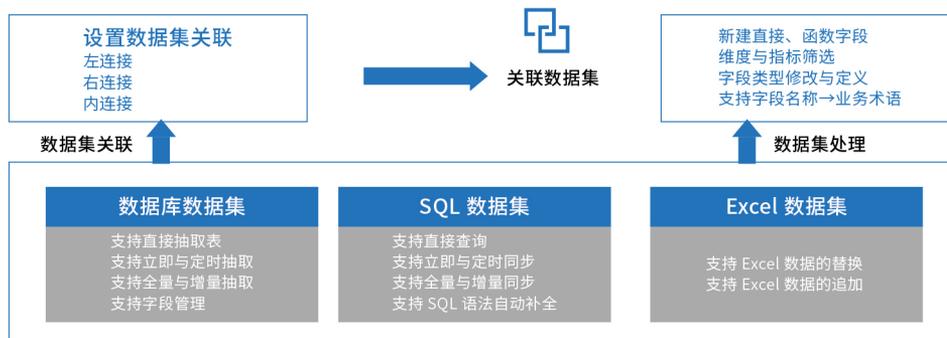


图12 DataEase 支持的常见数据来源

▶ 数据集

数据集相当于一组数据的实例，从对应的数据源中根据需求取用数据及组合数据，为下一步视图制作与数据分析提供数据基础。DataEase 目前支持的数据集创建方式有数据库数据集、SQL 数据集、Excel 数据集、自定义数据集、关联数据集、API 数据集共六种。

其中：数据库数据集指直接选择数据库中某一表作为数据集；SQL 数据集指可通过编写 SQL 语句将其查询结果作为数据集；Excel 数据集指通过导入本地 Excel 数据文件生成数据集；自定义数据集可根据个人对字段的需求，从现有数据集中灵活选择字段组成的新数据集；关联数据集指可在现有数据集基础上设置连接字段，拼接多个数据集，并选择需要的字段后形成的新数据集。



► 视图

视图是仪表板的组成元素之一，通过各类图表形式更直观地展示数据集中的内容，帮助用户更好地理解及表达数据。

视图目前支持两种图表库：蚂蚁数据可视化 AntV 和 Apache ECharts。支持的视图类型有：表格类型（汇总表、明细表、透视表），指标类型（指标卡、仪表盘），趋势类型（基础折线图、堆叠折线图、组合图），比较类型（基础柱状图、堆叠柱状图、横向柱状图、横向堆叠柱状图），分布类型（饼图、南丁格尔玫瑰图、雷达图、矩形树图），关系型（散点图、漏斗图），空间位置（地图）等。



图14 DataEase 核心概念：视图

► 仪表板

用户可以在仪表板的画布上将各类元素（视图、文字、图形、时间控件、过滤控件、选项卡、图片、视频等）用拖拉拽的方式组合起来，从而达到展示的目的。同时，完成的仪表板可以通过系统内授权或生成共享链接快速分享给他人，最终用户可以通过 Web、大屏、移动端等多种途径进行实时查阅。



图15 DataEase 核心概念：仪表板

4.2 DataEase 的功能列表

功能模块	功能名称	功能描述
首页	欢迎导航页	DataEase 常用文档、视频教程、更新日志等内容导航；
仪表盘	仪表盘管理	支持仪表盘的新建、重命名、删除、复制、移动、搜索等；
		支持仪表盘分组的新建、重命名、删除、移动等；
		支持以树状形式展示仪表盘分组；
	仪表盘制作	支持在线编辑仪表盘；
		支持仪表盘中添加多种组件，例如：视图 / 时间组件 / 文本组件 / 数字组件 / 样式组件 / 图片 / 视频 / Tab 组件 / 链接等；
		支持动态设置日期组件的默认值等；
		支持通过一个过滤组件，过滤多个视图（视图数据来自多个数据集）；
		支持组件样式设置，例如图形属性、组件样式等；
		支持仪表盘中视图的下钻；
		支持仪表盘中视图间的联动；
		支持仪表盘中各组件背景图片及边框的设置；
		支持仪表盘跳转，例如跳转至系统内其他仪表盘、外部链接；
		支持仪表盘背景、组件间隙、刷新时间、展示数据量等设置；
		支持一键切换仪表盘主题；
		支持仪表盘导出为 PDF 文件；
		支持撤销、重做、清空画布内容；
		支持仪表盘的全屏预览；
		支持仪表板的收藏；
		支持默认仪表板的设置；
		支持仪表盘模板的导出；
	支持仪表盘模版保存；	
	仪表盘共享	支持按组织 / 角色 / 用户分享，查阅分享记录等；
		支持创建公共链接，外部用户可通过密码访问仪表盘，同时分享链接支持设置过期时间；
	视图制作	支持通过简单的拖拉操作，制作视图；
		支持多种图表类型，明细表 / 汇总表 / 透视表 / 指标卡 / 基础柱状图 / 堆叠柱状图 / 横向柱状图 / 横向堆叠柱状图 / 基础折线图 / 堆叠折线图 / 饼图 / 南丁格尔玫瑰图 / 漏斗图 / 雷达图 / 仪表盘 / 中国地图 / 气泡地图 / 散点图 / 气泡图 / 矩形树图 / 组合图 / 水波图 / 瀑布图 / 词云图等；
		支持对柱状图、折线图等有纵坐标的图表，设置纵坐标起始和结束值范围；
		支持视图数据集的切换；
		支持视图数据集的编辑；
		支持选择视图的样式优先级；
		支持选择图表的排序方式，根据维度、指标升序、降序进行展示；
		支持视图的下钻和上卷；
		支持指标的多种汇总计算方式，例如求和、平均、最大值、最小值等；
		支持指标的高级计算，例如同比、环比等；
支持对图表类型的图形属性进行设置；		
支持对图表类型的组件样式进行设置；		
支持为视图添加过滤器，通过组合条件筛选数据；		

功能模块	功能名称	功能描述
数据集	数据集管理	支持添加多种类型的数据集，数据库数据集 / SQL 数据集 / Excel 数据集 / 自定义数据集 / 关联数据集 / API 数据集；
		支持 Excel 数据集数据的替换、追加；
		支持数据集的添加、移动、重命名、删除、预览等；
		数据库数据集和 SQL 数据集支持直连和定时同步两种连接方式；
		定时同步类型数据集，支持全量更新和增量更新两种方式；
		支持创建定时任务，以此控制数据集的更新；
		支持定时更新任务的查看；
		支持对数据集的字段类型 / 字段名 / 展示字段进行设置；
		支持数据集的维度、指标间的互相切换；
		支持自定义计算字段（内置常用计算函数支持）；
支持创建数据集间的关联关系（左连接、右连接、内连接）；		
数据源	数据源管理	支持多种数据源，例如：多 Sheet 页的 Excel 文件，MySQL / Oracle / SQL Server / PostgreSQL / Elasticsearch / ClickHouse / MongoDB / AWS RedShift / MariaDB / Doris / Hive / DB2 数据库 / API 数据源等；
		支持数据源的新建、编辑、删除等；
		支持数据源的高级设置，例如最大连接数、最小连接数、连接超时时间等；
		支持数据源有效性校验；
系统管理	用户管理	支持用户的新建、编辑、删除、修改密码、启用、禁用、搜索等；
		支持给用户分配组织；
		支持给用户分配角色；
	角色管理	支持角色的新建、编辑、删除、搜索等；
	组织管理	支持组织的新建、编辑、删除、搜索、排序、移动等；
	权限管理	支持从组织、角色、用户维度（组织架构维度）进行使用、管理、授权等形式的权限控制；
		支持从数据源、数据集、仪表盘（资源维度）进行使用、管理、授权等形式的权限控制；
		支持菜单和操作层面的权限控制；
		支持数据集的行级权限控制；
	支持数据集的列级权限控制；	
	显示设置	支持头部系统 Logo、登录页 Logo、登录页图片、登录页标题、系统名称等显示设置；
	主题设置	支持两种默认主题；
		支持自定义主题的新建、编辑、删除等；
	支持对主题进行基础配色、字体配色、边框配色、背景配色等多属性的设置；	
	LDAP 设置	支持 LDAP 认证对接；
	单点登录	支持单点登录系统对接；
	集成与扩展	提供完善的 API 接口及文档；
模板管理	支持系统模板和用户模板；	
	支持模板的分类、导入、重命名、删除、搜索等；	
消息管理	支持系统常见消息的通知；	
	支持消息的接收配置；	
	支持消息状态标记；	
支持已读消息的删除；		
任务管理	支持对平台定时任务进行全生命周期管理；	
	支持定时报告，可以定时以邮件形式发送仪表板报告。	

4.3 DataEase 的核心优势

4.3.1 简单易用的操作方式

DataEase 最重要的特性之一是“人人可用”，这是 DataEase 在每一次开发迭代与功能设计时的指导方向，更是最终的目标。DataEase 的“人人可用”具体体现在以下几个方面：

- ▶ 安装部署上，DataEase 支持快速获取在线或离线安装包，提供一键安装命令。软件升级操作同样简单统一，可一键自动升级至最新版本；
- ▶ 功能上，DataEase 在视图及仪表盘等操作上以拖拉拽方式为主，业务人员无需具备专业性技能即可完成各类数据的视图及仪表盘制作；
- ▶ 反馈与支持上，DataEase 的交流沟通渠道有 GitHub Issues、企业微信交流群、咨询电话和邮箱，后期还会有相关的技术问答社区等。多元化的开发者与使用者的交互途径一方面可以帮助用户更好地使用 DataEase 来实现价值，另一方面产品可以听取广大用户的声音实现进一步的优化；
- ▶ 学习资源上，DataEase 具备丰富完善的开源线上指导文档，同时有技巧操作类、仪表盘分享类等视频资源，每月迭代后会通过直播平台对外线上演示新功能，并解答用户在使用过程中遇到的问题。此外，DataEase 很快将会推出专属的学习班课程，为广大用户提供更加专业的学习培训服务。

4.3.2 开源开放的运营模式

DataEase 开源数据可视化分析平台作为飞致云旗下的开源产品之一，与 JumpServer 开源堡垒机、MeterSphere 开源持续测试平台的理念一致，长期坚持开源运营模式，按月进行版本迭代。自 2021 年 6 月 28 日发布以来，DataEase 收获了一份还不错的成绩单，并在过去几个月中多次跻身 GitHub Java 项目趋势榜前列。截至 2022 年 3 月 18 日，DataEase 开源项目在代码托管平台 GitHub 上的数据表现如下：

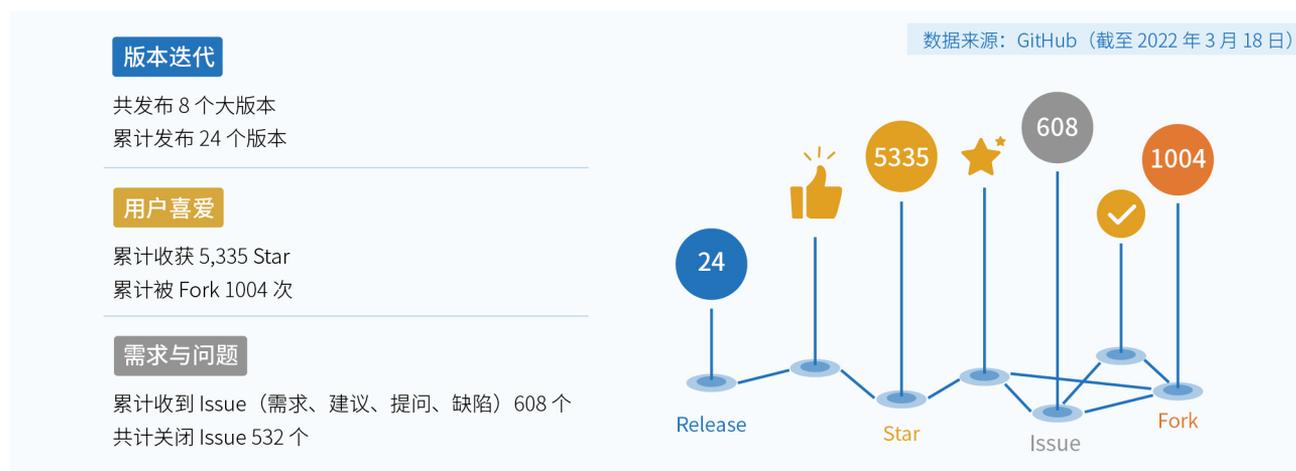


图 16 DataEase GitHub 数据

从以下趋势图也可看出，DataEase 在 GitHub 代码托管平台的 Star 数量一直处于稳定增长的状态。

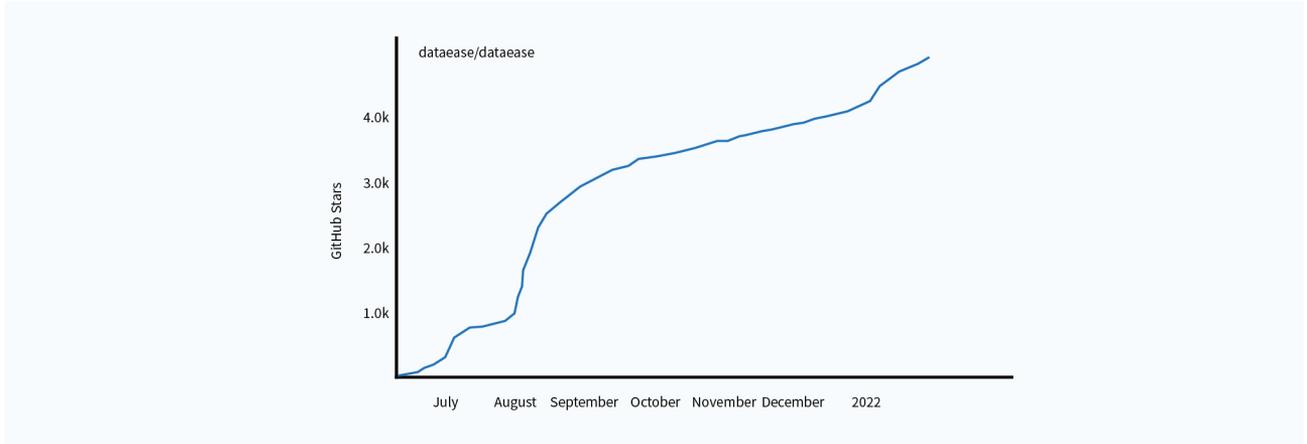


图17 DataEase GitHub Star 趋势图

DataEase 还具备完善的开源线上文档、企业微信交流群等渠道，能够为广大开源用户快速上手使用提供更多的支持。其中，线上文档包括系统功能架构、基本操作说明、安装部署指导、本地二次开发说明、仪表板制作示例、产品使用技巧等内容。企业微信交流群的数量也在半年多的时间内增长到 16 个，DataEase 开源项目组一直活跃其中，帮助用户解决使用过程中的实际问题，同时倾听用户建议，将来自一线用户的反馈注入到下一版本的研发工作之中。

值得一提的是，2021 年 12 月，知名开源技术社区 OSCHINA（开源中国）公布了“2021 年度 OSCHINA 优秀开源技术团队”获奖名单。FIT2CLOUD 飞致云凭借在开源软件研发和开源社区运营方面的积极表现，荣膺“2021 年度 OSCHINA 优秀开源技术团队”奖项。这正是对 FIT2CLOUD 飞致云近几年大力推广、运营多款开源产品的有力肯定。获得了用户和社会的广泛认可后，DataEase 将继续践行开源的理念，为企业与个人提供更好的数据分析平台。

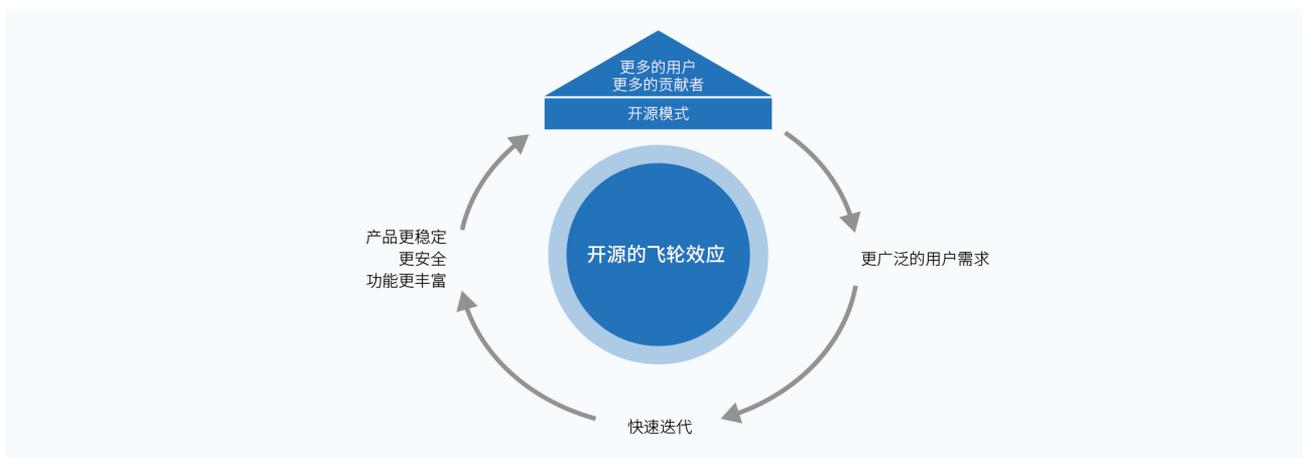


图18 开源的飞轮效应

4.3.3 大数据下的秒级响应

数据可视化分析工具的另外一个核心能力指标，在于其可以支持的数据量级以及在大数据场景下的性能表现，因此 DataEase 在架构设计以及组件选型时也仔细考量了这一因素。

架构设计上，DataEase 采用容器化部署的模式，简化了部署的复杂性，同时支持各组件灵活进行分布式部署与水平扩展。组件选型上，DataEase 选择了 Apache Doris 作为本地模式下的 OLAP 数据引擎。Doris 最早由百度自主研发，在其自身业务需求下应运而生，后期开源并捐献给 Apache 基金会。Doris 属于 MPP 分析型数据库产品，仅需秒级甚至毫秒级响应时间即可获得海量数据下的查询结果，有效地支持实时数据分析、交互式数据分析等多种需求场景。Doris 的分布式架构非常简洁，易于运维，并且可以支持 10 PB 以上的超大数据集。



图19 Apache Doris 核心特性

4.3.4 安全分享

正如上文所述，仪表板的共享与分享同样是可视化分析中的重要一环。DataEase 注重仪表板分享功能的建设，包括分享操作的便捷性及分享过程的安全性。

► 以创建公共链接的方式分享仪表板

DataEase 支持创建公共分享链接，同时支持为公共链接设置访问密码及链接有效期。用户也可选择不为仪表板设置访问密码，作为公开内容分享，其他任何用户可免登录直接使用该链接查看仪表板内容，或是将仪表板嵌入到其他系统进行辅助展示。用户若设置访问密码，则可以有效保障仪表板的数据安全，对访问人群进行选择 and 限制。同时，设置链接有效期可以控制公共链接的受访期限，进一步增强安全性。此外，仪表板所使用的数据集若有行列权限限制，则分享者的仪表板数据权限也仅限于其所具备的数据权限。



图20 DataEase 公共链接分享及访问模式

► DataEase 系统内部仪表板分享

系统内部仪表板分享是指被分享者属于 DataEase 内部用户，需登录验证自己账号后进行查看，这一模式具备更高的安全性。系统内部分享支持基于组织、角色以及用户维度进行操作。分享者可在“分享”目录下查看自己历史分享过的仪表板以及分享对象，并可以随时取消分享。

5. DataEase 仪表板案例集

以下我们将分享自 DataEase 项目发布以来，用户投稿的一些仪表板案例。更多详细内容及仪表板分享可关注“DataEase 开源数据可视化分析平台”官方微信公众号获取。

▶ JumpServer 运维数据仪表板

JumpServer 是全球首款开源的堡垒机，使用 GPLV3 开源协议，是符合 4A 规范的运维安全审计系统。用户所在公司一直使用 JumpServer 加强运维安全管控并进行定期的运维审计。用户通过 DataEase 数据可视化工具丰富 JumpServer 的审计报表，并且对一些审计数据进行自定义展示。

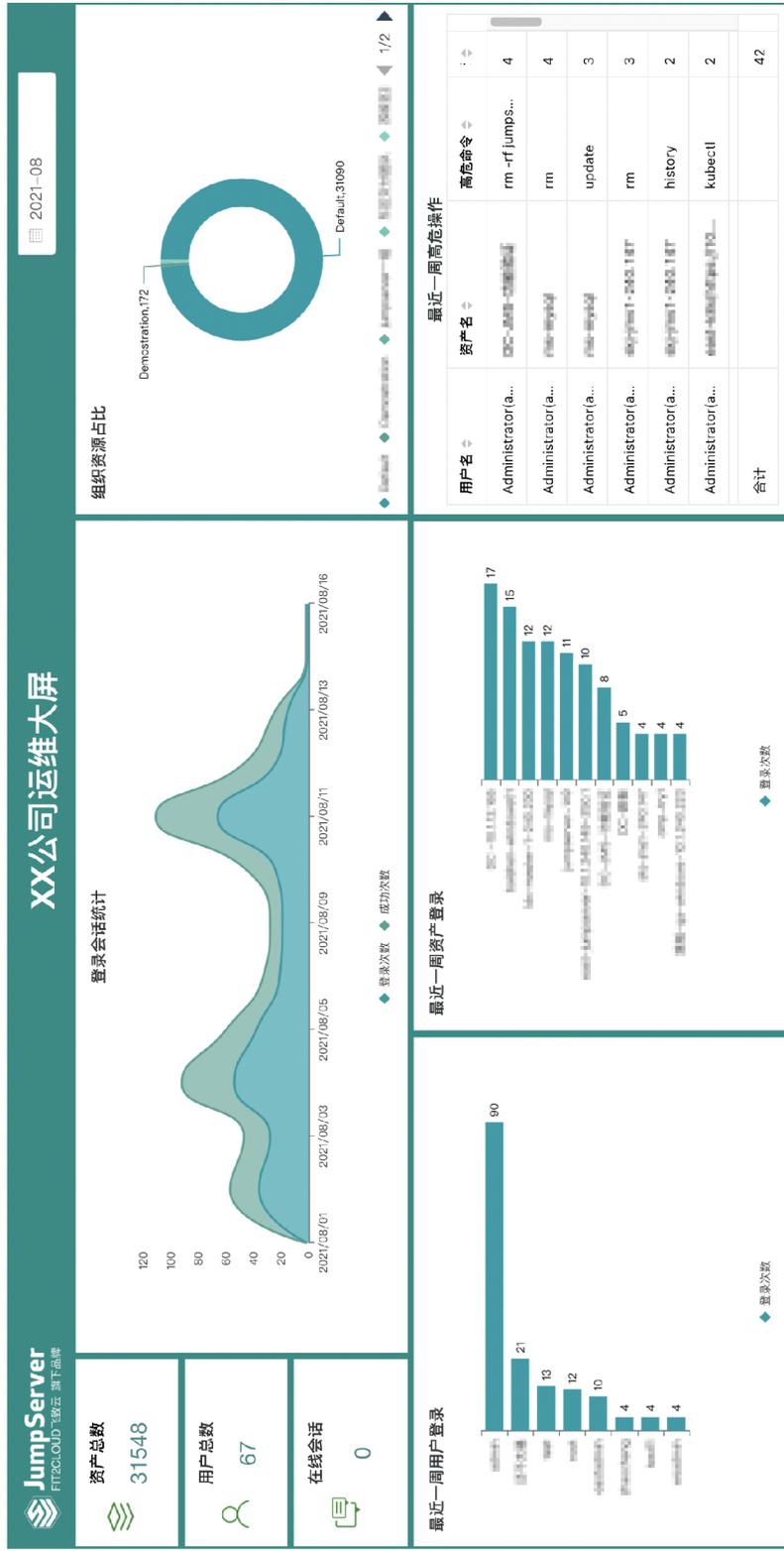


图21 JumpServer 运维数据仪表板

► 居民消费价格趋势分析仪表盘

该用户使用 DataEase 制作了 2021 年 8 月居民消费价格数据的仪表盘。用户通过国家统计局官网获取了“2021 年 8 月份居民消费价格”数据，并导入到 DataEase 平台，再进一步使用拖拉拽的方式完成后续制作。最终展现了 2021 年 8 月全国居民消费价格同比 / 环比涨跌幅、各类商品及服务价格同比 / 环比变动情况，以及居民消费价格等主要数据。

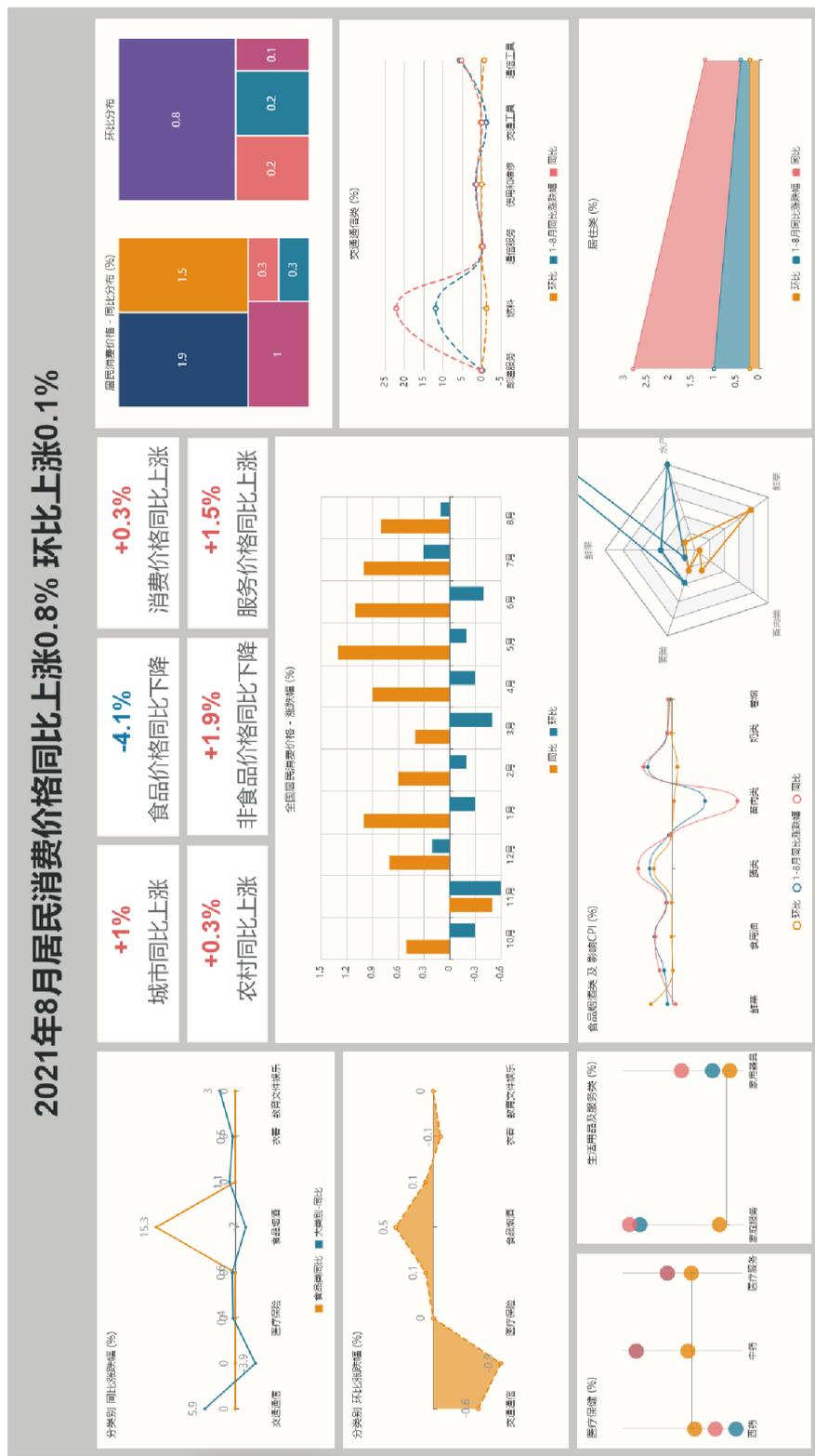


图22 2021 年 8 月居民消费价格趋势分析仪表盘

► MeterSphere 测试用例仪表盘

MeterSphere 是一站式的开源持续测试平台，涵盖了测试跟踪、接口测试、性能测试、团队协作等功能。用户所在企业于 2020 年开始使用 MeterSphere，用于进行测试用例管理并对接口和性能进行测试。随着业务发展以及对 MeterSphere 的深度使用，该公司需要更多维度、更丰富好看的大屏进行系统数据展示，并要求仪表盘可以快速分享给他人。借助 DataEase 开源工具，用户制作了测试用例数据实时概览的仪表盘。

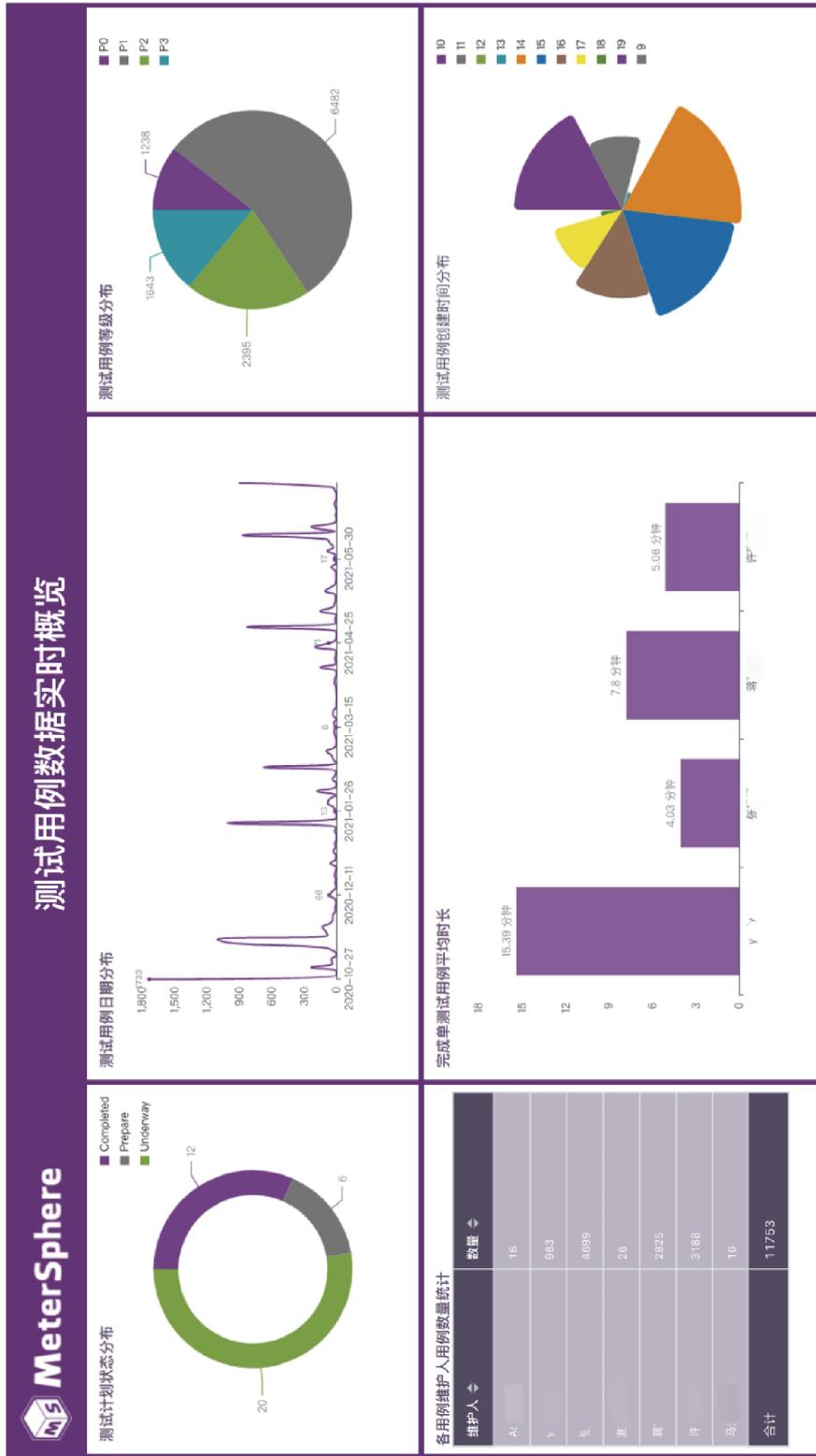


图23 测试用例仪表盘

6. 参考资料

- [1] 中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要 [Z]. 国务院,2021.
- [2] 开源生态白皮书 [R]. 中国通信院,2021.
- [3]GitHub.The 2021 State of the Octoverse[R].<https://octoverse.github.com/>,2021.
- [4] 陈为,沈则潜.数据可视化 [M]. 电子工业出版社,2013.
- [5]Cole Nussbaumer Knaflic.用数据讲故事 [M]. 人民邮电出版社,2017.
- [6]2021 年影响中国开源未来的十大热点 [R].OSCHINA,2021.
- [7] 邱南森 (Nathan Yau). 数据之美 [M]. 中国人民大学出版社,2014.
- [8] 中国数字经济发展白皮书 [R]. 中国通信院,2021.
- [9] 中国人民银行、中央网信办等五部门关于规范金融业开源技术应用与发展的意见 [R]. http://www.cac.gov.cn/2021-10/27/c_1636928705274546.htm,2021.
- [10]The State of Enterprise Open Source[R].RedHat,2021.



FIT2CLOUD 飞致云

联系电话: 400-052-0755

网站地址: www.dataease.io

电子邮件: support@fit2cloud.com

Copyright © 2022-2024 FIT2CLOUD, Inc. All rights reserved.



DataEase 微信公众号



DataEase 交流微信群