ISCAS 中国科学院软件研究所学术年会'2022
暨计算机科学国家重点实验室开放周

重大示范
工程/项目

开源软件供应链重大基础设施

中国科学院软件研究所、中科南京软件技术研究院 联合攻关团队

崔星 电话：13051316652 邮箱： cuixing@iscas.ac.cn

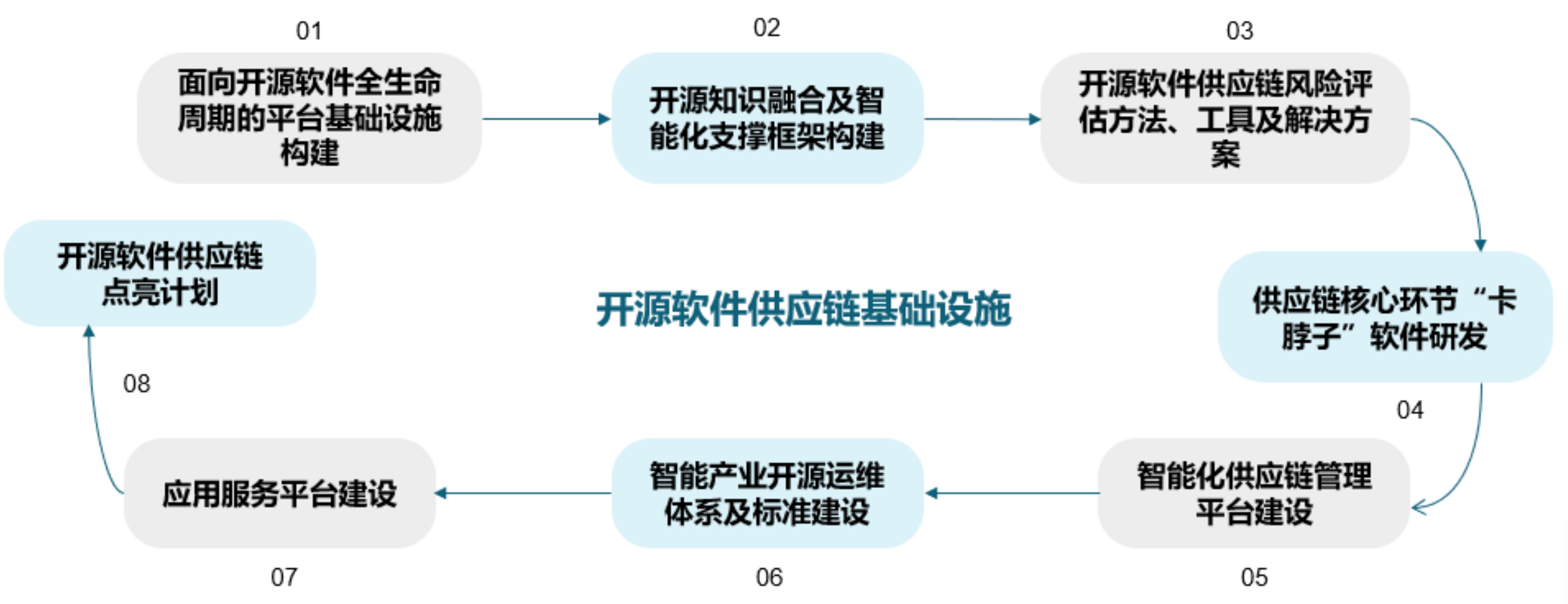
项目背景

全球范围内，开源软件已经成为了基础设施的核心要素，是构成操作系统、数据库等重要“卡脖子”基础软件的核心“元器件”。能否为设备、系统、产业和行业提供高质量的、高可靠的、可持续演进的开源软件供应，关系到国内当前和未来 IT 科研、产品与生态的核心竞争力，甚至是“生命线”。然而，近年来，国内开源软件供应链“卡脖子”事件频频发生，开源软件产业仍面临着不少根本问题。为了贯彻落实发展开源软件的国家战略，实现开源软件的可靠供应，需要尽快打造核心基础设施支撑，解决基础软件核心技术受制于人的问题。

开源软件供应链重大基础设施旨在应对开源软件供应存在的风险，突破基础软件领域关键核心技术，建设国内首个开源软件采集存储、开发测试、集成发布、升级运维一体化设施，为国内关键设备、系统、行业和产业提供高质量、低风险的开源软件供应链，实现开源软件可靠供应，打破国外巨头在软件行业价值链垄断。进而打造服务全球的开源代码知识图谱和开源软件供应链体系，保障我国的软件供给安全和产业创新发展。

项目内容

开源软件供应链重大基础设施为软件科学研究提供所需的开源代码大数据，并提供高度结构化的数据组织形式，从而支撑软件工程的智能化，为机器智能编程等信息技术前沿领域奠定基础。


开源软件供应链基础设施
01 面向开源软件全生命周期的平台基础设施构建
02 开源知识融合及智能化支撑框架构建
03 开源软件供应链风险评估方法、工具及解决方案
04 供应链核心环节“卡脖子”软件研发
05 智能化供应链管理平台建设
06 智能产业开源运维体系及标准建设
07 应用服务平台建设
08 开源软件供应链点亮计划

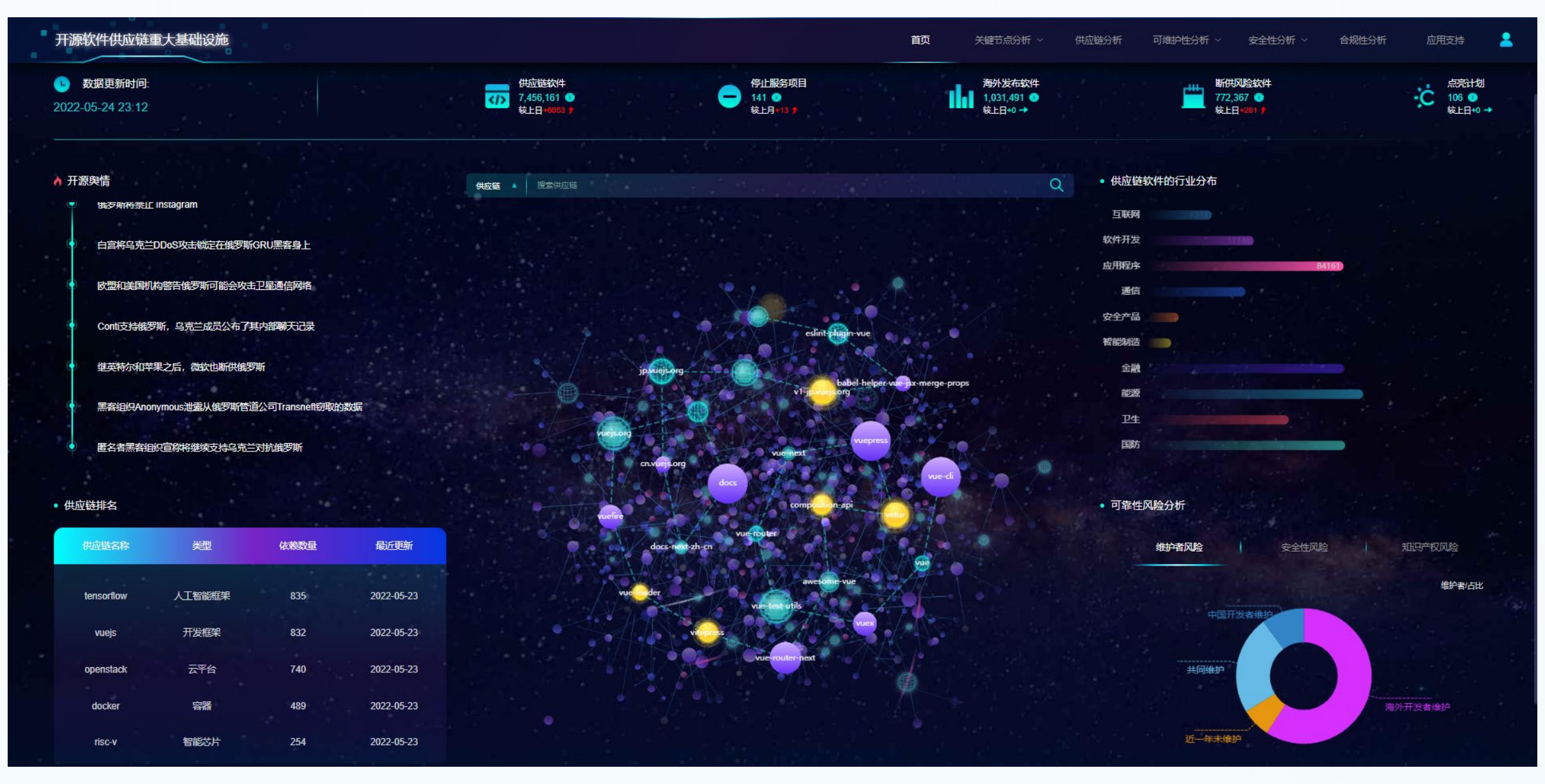
大规模开源软件数据集：面向海量开源软件构建开源软件知识图谱，以开源软件知识图谱为核心，辅以链路预测、事件抽取、基于神经网络的风险分析器、开源软件漏洞分析等技术手段，为分析开源软件可靠供应链提供有力支撑。

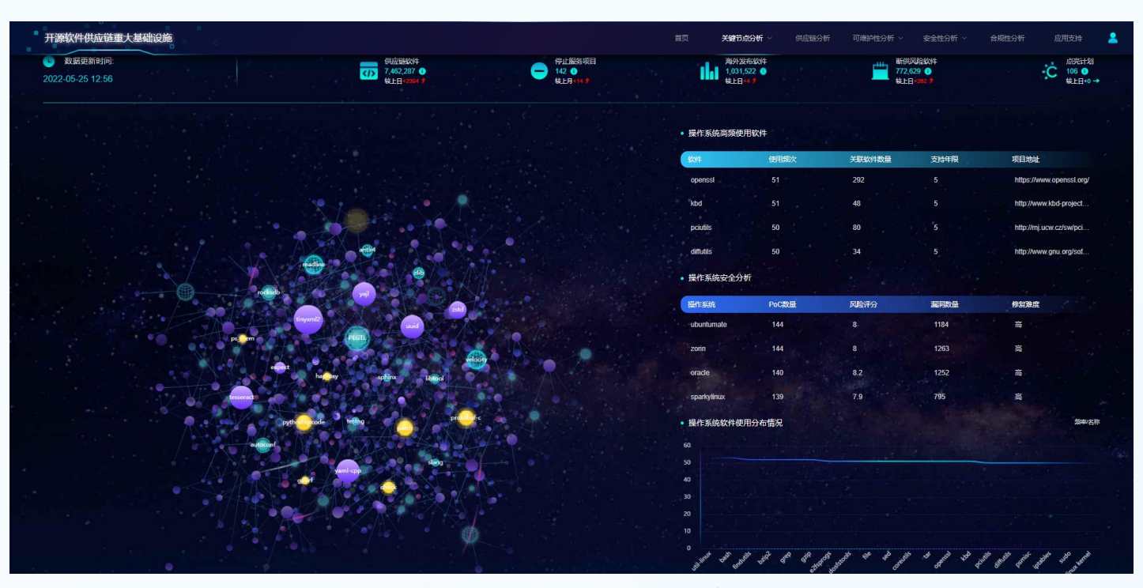
实体类型	总数
软件	7063136
漏洞	262067
存储仓库	1407996
组织	127275
开发者	3692033
地理位置	2243

关系类型	总数
依赖关系	39617325
漏洞影响关系	19756614
仓储关系	4719060
维护关系	1095966
贡献关系	9177351
位置关系	1113159

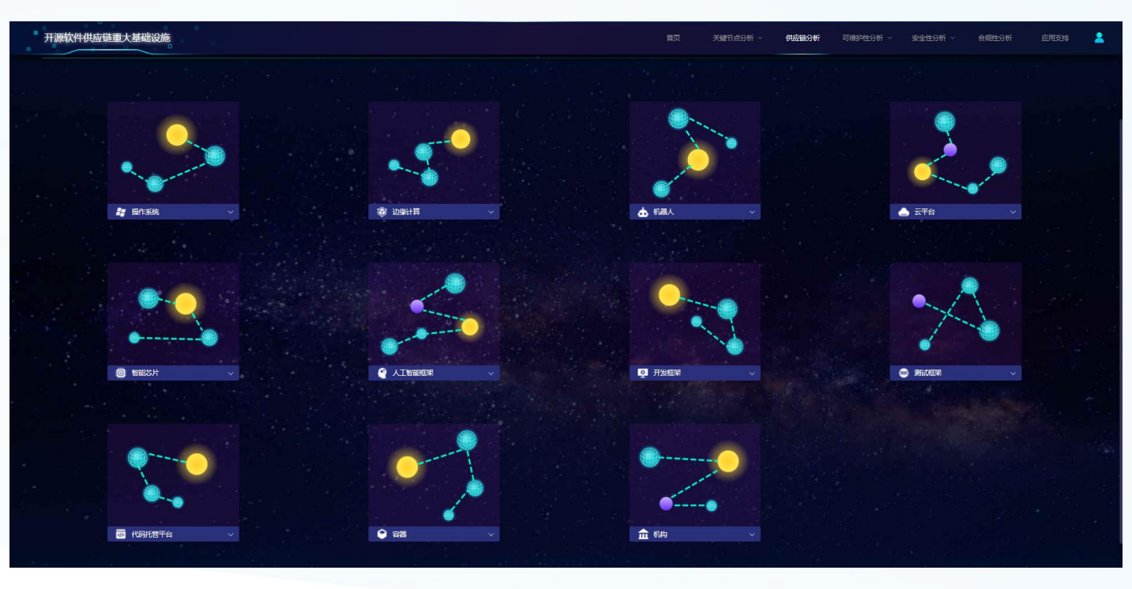
目前，软件图谱中共有实体31种，实体与实体间关系56种，实体与属性之间关系787种。实体数量达1523万，关系数量超过1.7亿条

重大基础设施应用平台：开源软件供应链重大基础设施平台集成了关键节点分析、供应链分析、可维护性分析、安全性分析、合规性分析以及应用支持等功能，通过对软件数据进行分析，从多种维度挖掘出软件特征、软件与软件之间的关系和依赖，实现对软件的推理、预测、推荐等功能，构建开源软件可靠供应链。

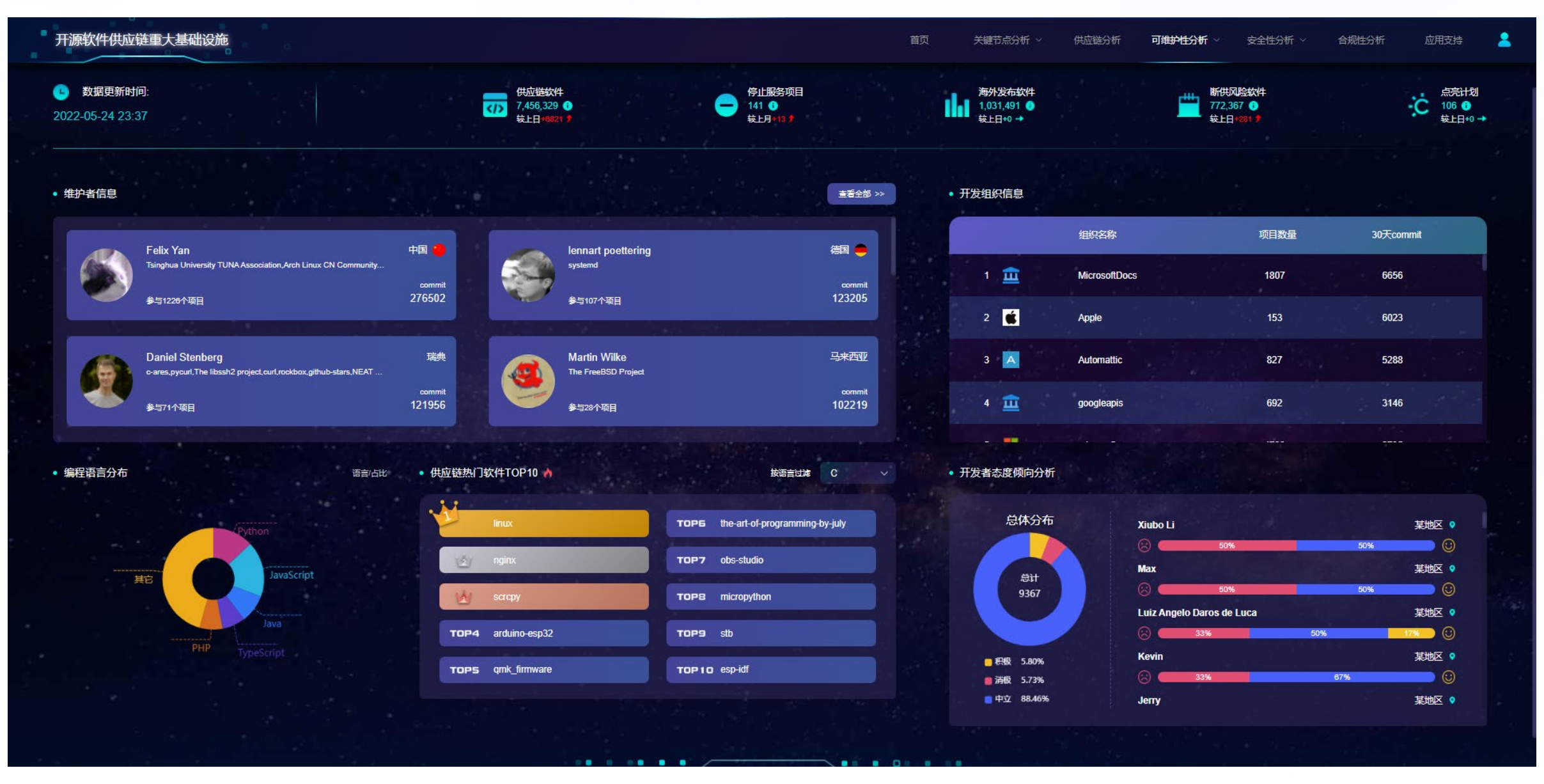

开源软件供应链重大基础设施平台


关键节点分析

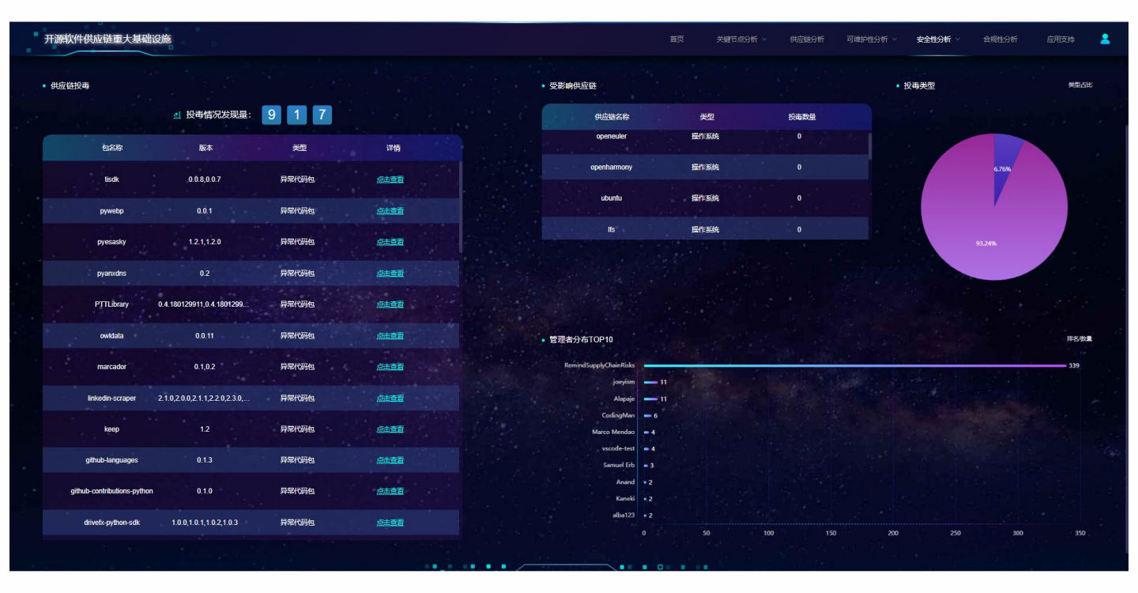
- 通过对操作系统高频使用软件进行分析，能够获取到操作系统中关键软件。通过判断已有操作系统的威胁风险，进一步为软件选型提供安全参考。


供应链分析

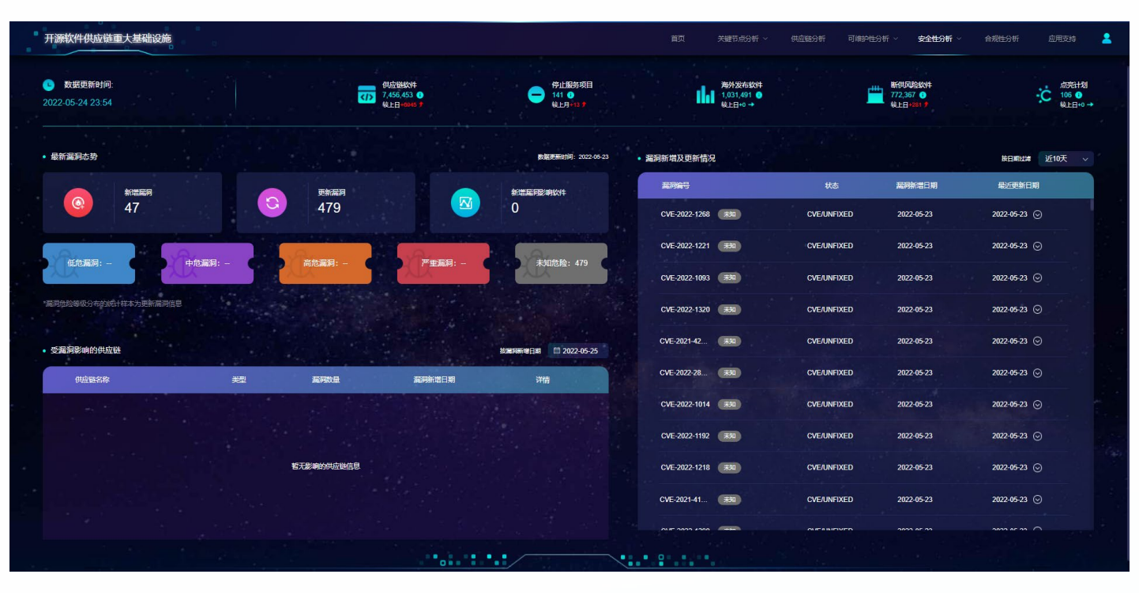
- 供应链分析集成了11种应用场景下的16条典型供应链，涉及了操作系统、智能芯片、云计算等领域。



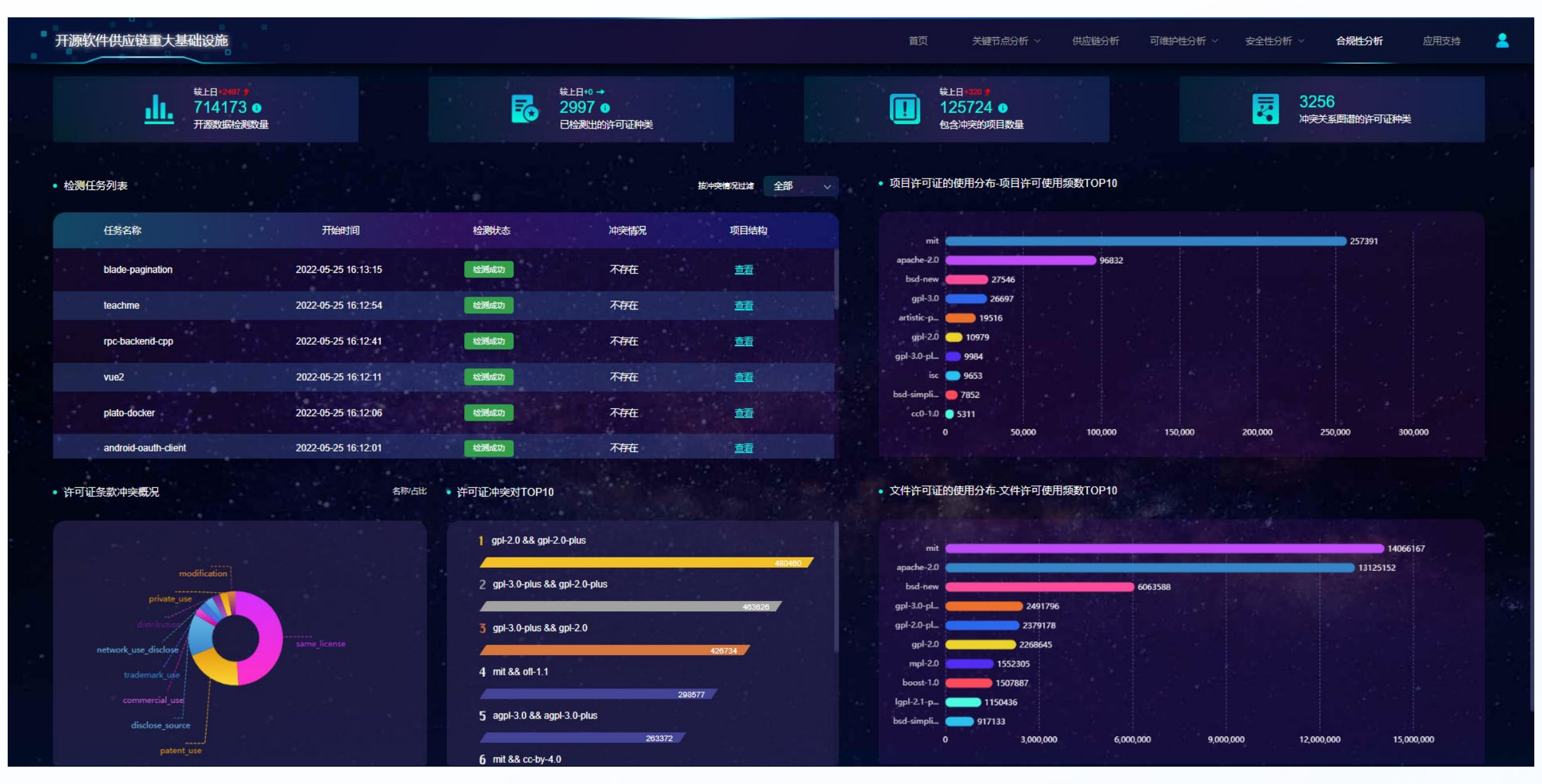
- 可维护性分析，包含对软件开发者、组织、软件热度、人员倾向等等多维度分析，为进一步判断软件的可靠性提供参考。


安全性分析—供应链投毒

- 恶意代码检测实现了对大量软件包的统计计算，及时发现风险，能够有效解决供应链软件投毒。


安全性分析—漏洞态势感知

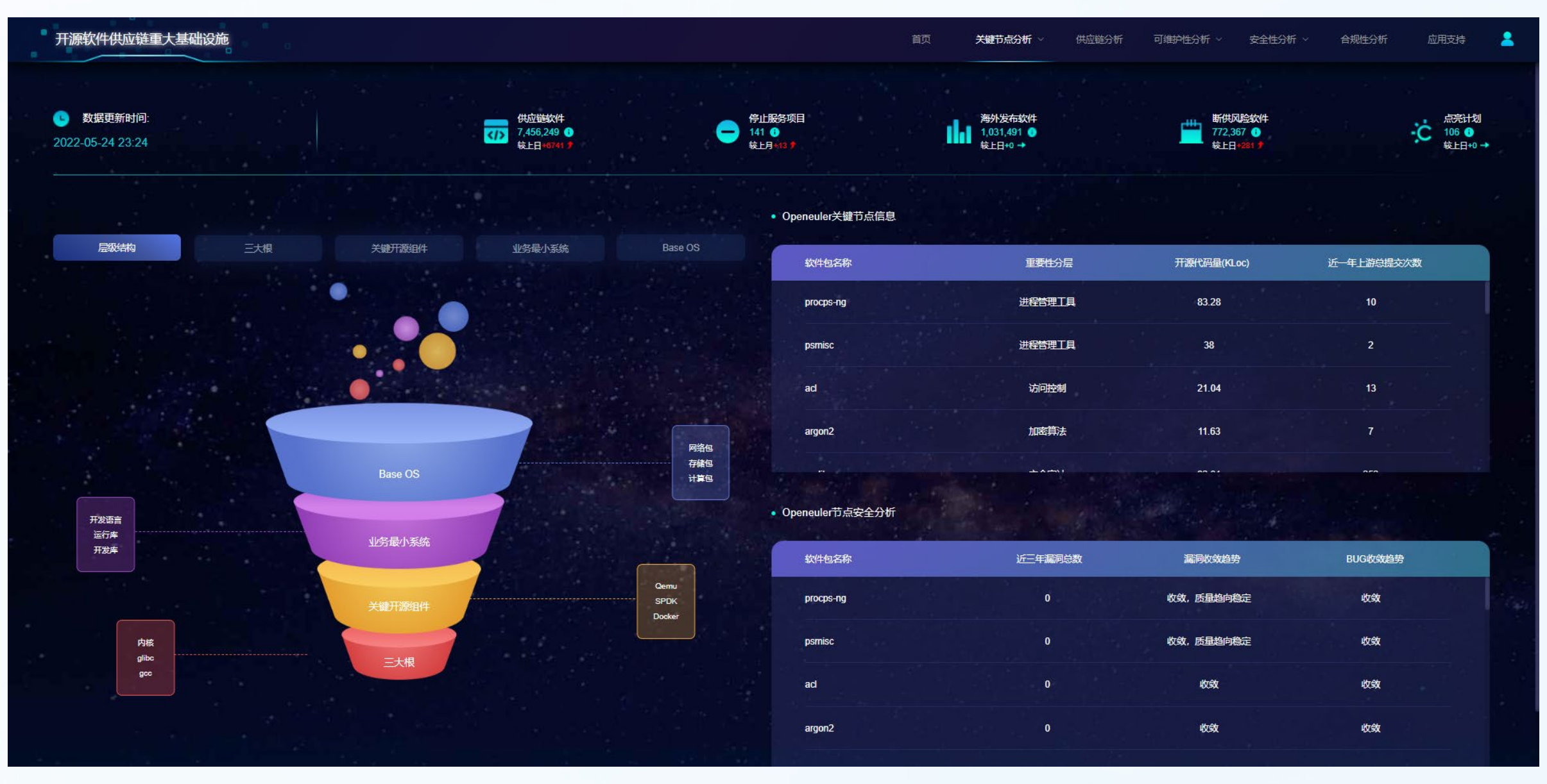
- 通过完整的漏洞生命周期跟踪实现风险闭环，对服务器/云/容器/边缘计算等多类计算环境进行实时监测。



- 合规性分析，通过信息检索、自然语言处理等技术建立知识库，识别开源许可证兼容性风险等知识产权问题。当前已自动化地处理了3259种开源许可证，并构建了许可证冲突关系知识库。

应用支撑

重大基础设施平台支撑openEuler基础软件的风险分析。基于软件图谱中8000+ openEuler 软件包，我们分析出其中的基础软件与部分关键软件。通过软件关联性分析，计算出与openEuler基础软件包具有较强关联关系的上下游软件。基于态势感知与影响范围推理，进一步判断出软件存在的漏洞风险。

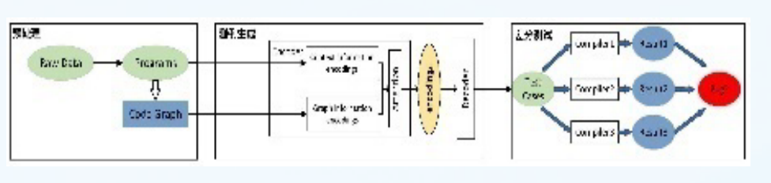

开源软件供应链重大基础设施中的openEuler分析

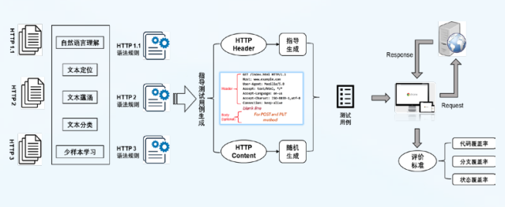
↑ 针对openEuler进行多维度知识计算

■ 操作系统及软件的风险评估

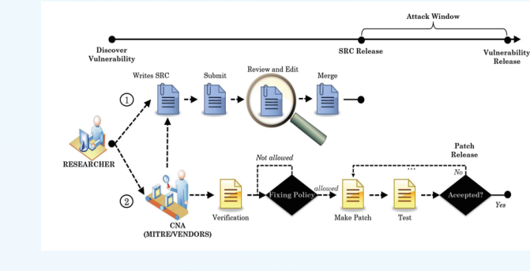
软件评估	合规性评估	安全评估
开发组织分析	任意编程语言的license检测	代码漏洞感知
维护者分析	3200+开源许可证知识库	漏洞态势感知
热门软件分析	许可证条款级别的冲突分析	供应链投毒分析

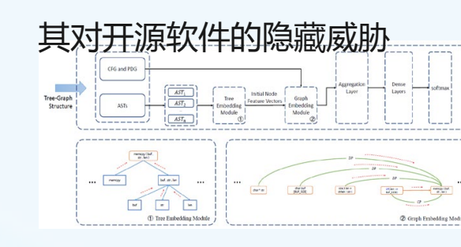
■ Fuzzing技术的研究


基于代码图生成的编译器模糊测试技术


基于RFC规范语法自动理解的HTTP协议模糊测试技术

■ 无CVE警示的漏洞感知


通过对潜在漏洞信息的数据挖掘，研究其对开源软件的隐藏威胁


通过基于深度学习的漏洞挖掘，实现漏洞关键特征准确识别

通过对操作系统、软件进行软件分析评估，开展漏洞感知、面向编译器及网络协议的模糊测试，进一步支撑openEuler基础软件的相关研究。

下一步发展目标

围绕开源软件供应链数据采集平台，应用服务平台、可视化展示平台、智能化供应链管理平台、以及开放软件包服务平台进行重点攻关，打造服务全球的开源代码知识图谱和开源软件供应链体系，保障我国软件供给安全和产业创新发展。