



2024 AI+研发数字峰会

AI+ Development Digital summit

AI驱动研发变革 促进企业降本增效

北京站 08/16-17



基于GenAI的混合云智能运维实践

周彩钦 联想



周彩钦

联想/混合云产品研发总监

联想混合云产品研发总监，负责联想xCloud产品的研发。在技术研发和团队管理方面拥有丰富的经验，成功带领团队开发了AIOps、云原生平台等前沿产品。通过这些创新，推动了联想智能运维转型，实现卓越运维。

目录

CONTENTS

1. 联想混合云智能运维发展历程
2. GenAI混合云智能运维架构
3. GenAI的运维应用场景
4. 实践经验与展望

PART 01

联想混合云智能运维发展历程

▶ 联想IT发展历程

1998-2004

信息化

ERP OA系统实施



数据

- 数据孤岛

运维

- 手工, 部分脚本化

应用开发

- 信息化系统逐步建立
- 传统瀑布式开发

基础架构

- 中国本土传统数据中心
- 传统基础架构

2005-2016
全球化+数字化

全球战略平台整合



数据

- 数据仓库
- 大数据平台

运维

- 自动化平台

应用开发

- 核心业务系统协同
- 双模式开发 (瀑布式+敏捷)

基础架构

- 全球数据中心
- 基础架构国产化替代
- 全面虚拟化+部分云化

2017- 未来
智能化

全球智能化平台建设



数据

- 数据治理
- 统一大数据平台
- 数据智能应用

运维

- 数据驱动的智能运维

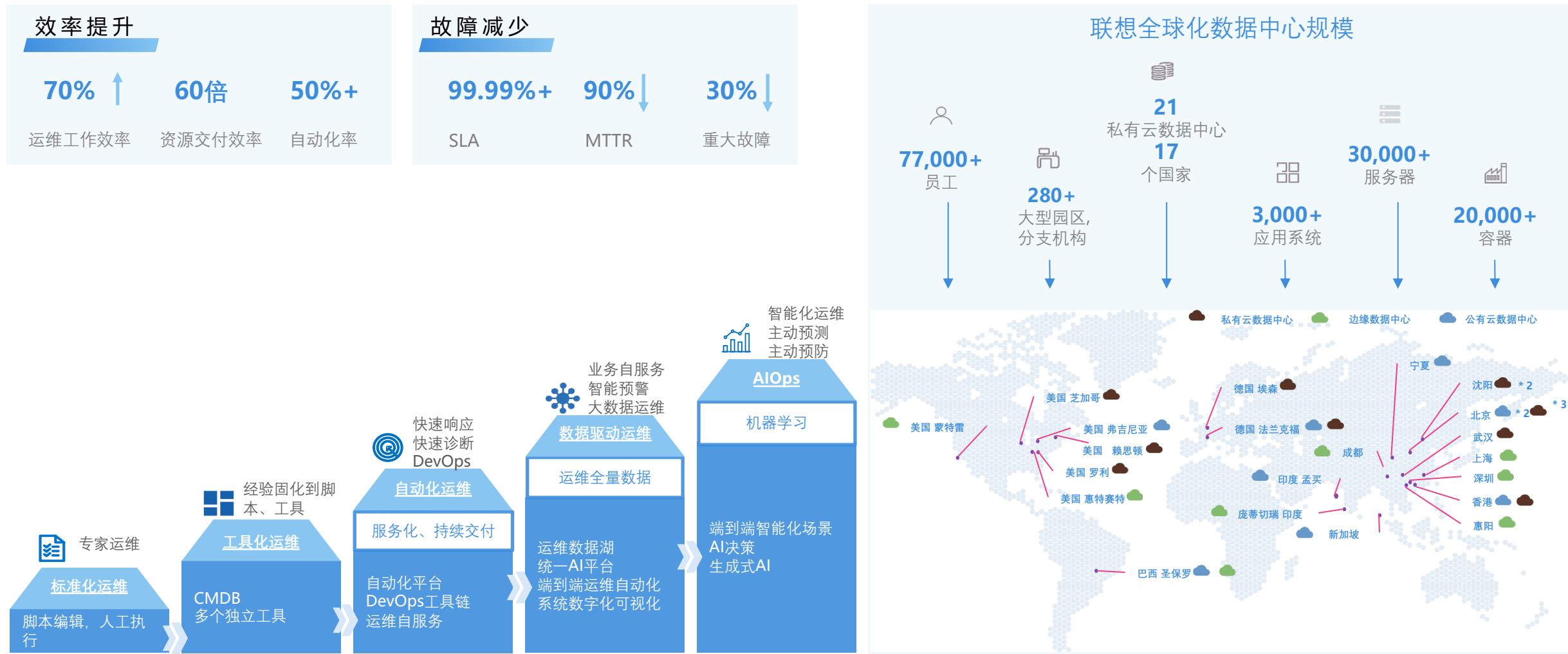
应用开发

- 云原生平台
- 应用的云原生改造
- 敏捷开发

基础架构

- 基础架构全面云化
- 混合云落地

联想智能运维发展历程



PART 02

GenAI混合云智能运维架构

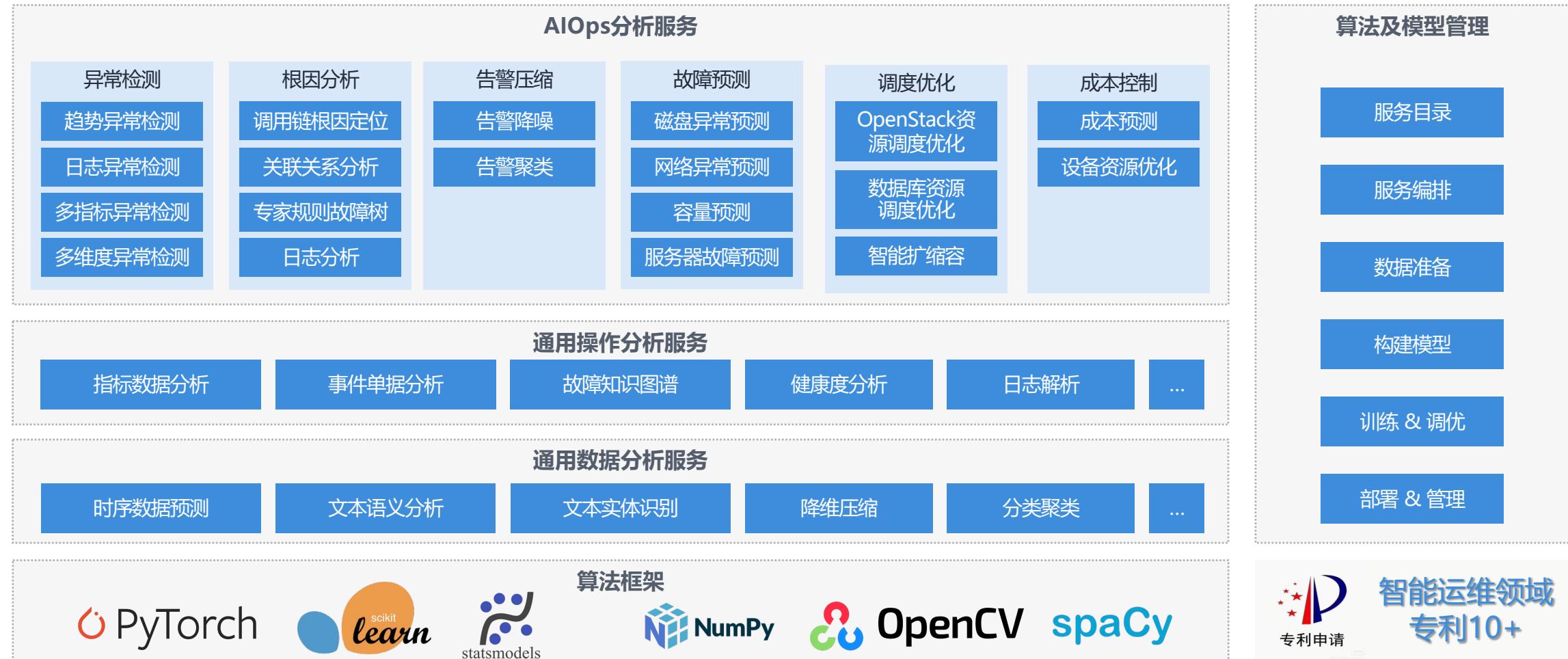
► GenAI混合云智能运维架构 - 联想xCloud AIOps

联想xCloud AIOps是企业级IT运维智能管理解决方案，以数据驱动的智能分析为脑，以可观测的监控告警为眼，以任务驱动的自动化平台为手脚，实现混合云环境端到端的智能管理。



▶ 智能分析平台设计

联想智能分析平台 (AI Analysis) 基于判别式AI技术，旨在为 xCloud AIOps 产品提供传统算法服务支持，涵盖从异常检测、根因分析到成本控制和资源优化的多个智能运维场景，助力企业提升IT运维效率和系统稳定性。



联想智小星 (xSpark) - 基于GenAI的IT运营智能体

联想智小星 (Lenovo xSpark) 是基于生成式AI的IT运营智能体，可嵌入或集成到 xCloud AIOps产品中。全面提升IT运维效率和用户交互体验。



业务价值

GenAI 赋能IT运营

依托xCloud产品生态，整合企业IT运营知识和工具，为企业提供一体化的智能运营解决方案

保障数据安全

- 支持内部部署，确保企业数据安全
- 与第三方系统开放式集成

提升运营效率

- 18% ↑ 根因分析效率提升
11% ↑ 运维研发效能
22% ↑ 任务执行效率

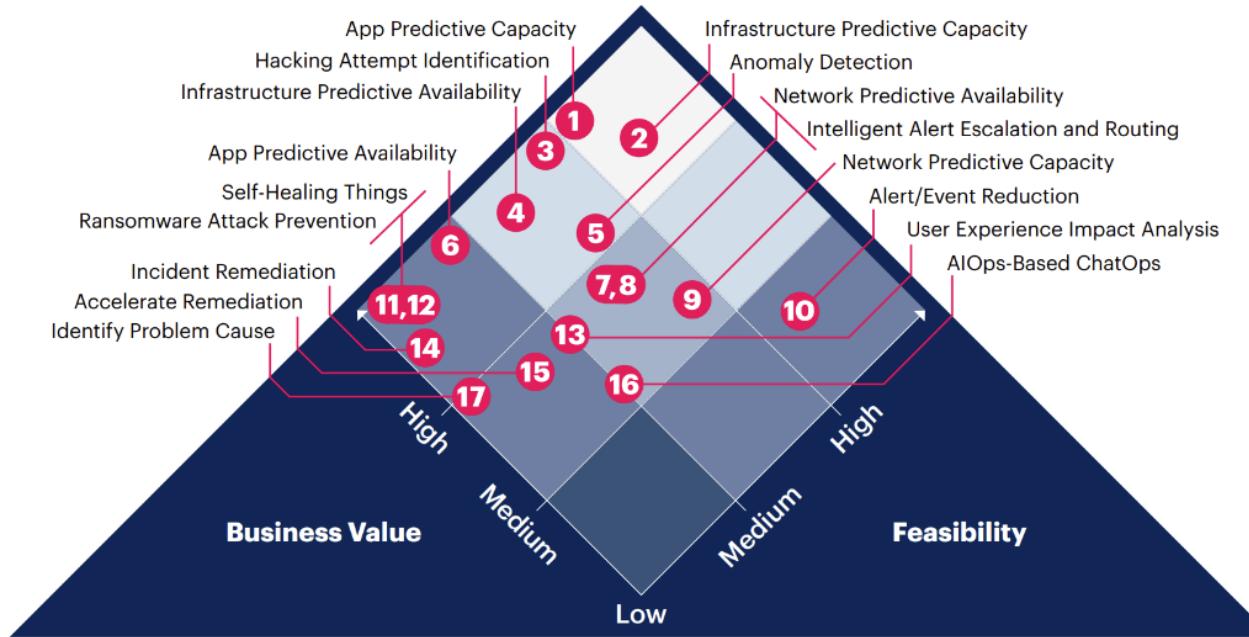
PART 03

GenAI的运维应用场景

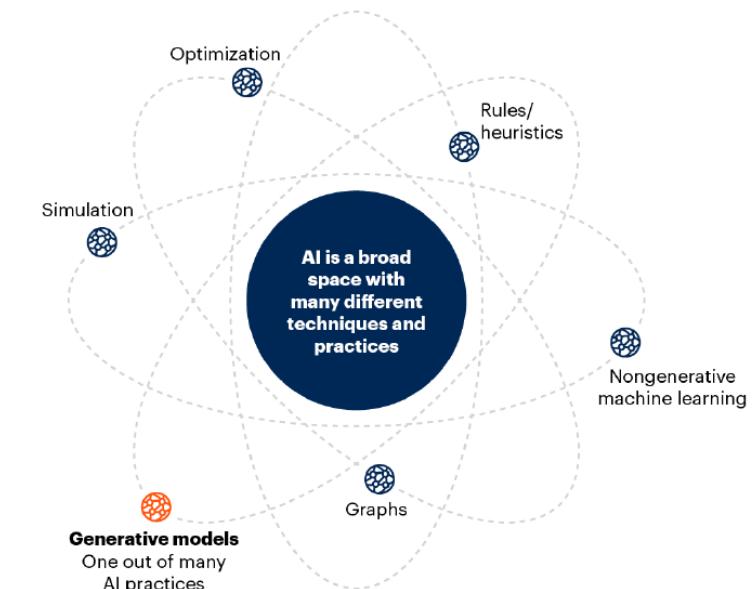
▶ 智能运维场景介绍 - Gartner

总结

- 运维场景从业务导向出发，场景分类和业务价值没有本质改变；
- 结合GenAI技术，应更多关注对技术可行性评估带来的改变和技术实现路线；
- GenAI is not a silver bullet. many business problems will require a combination of different AI techniques.
(Gartner)



AI Does Not Revolve Around GenAI



Source: Gartner
806238_C

Gartner

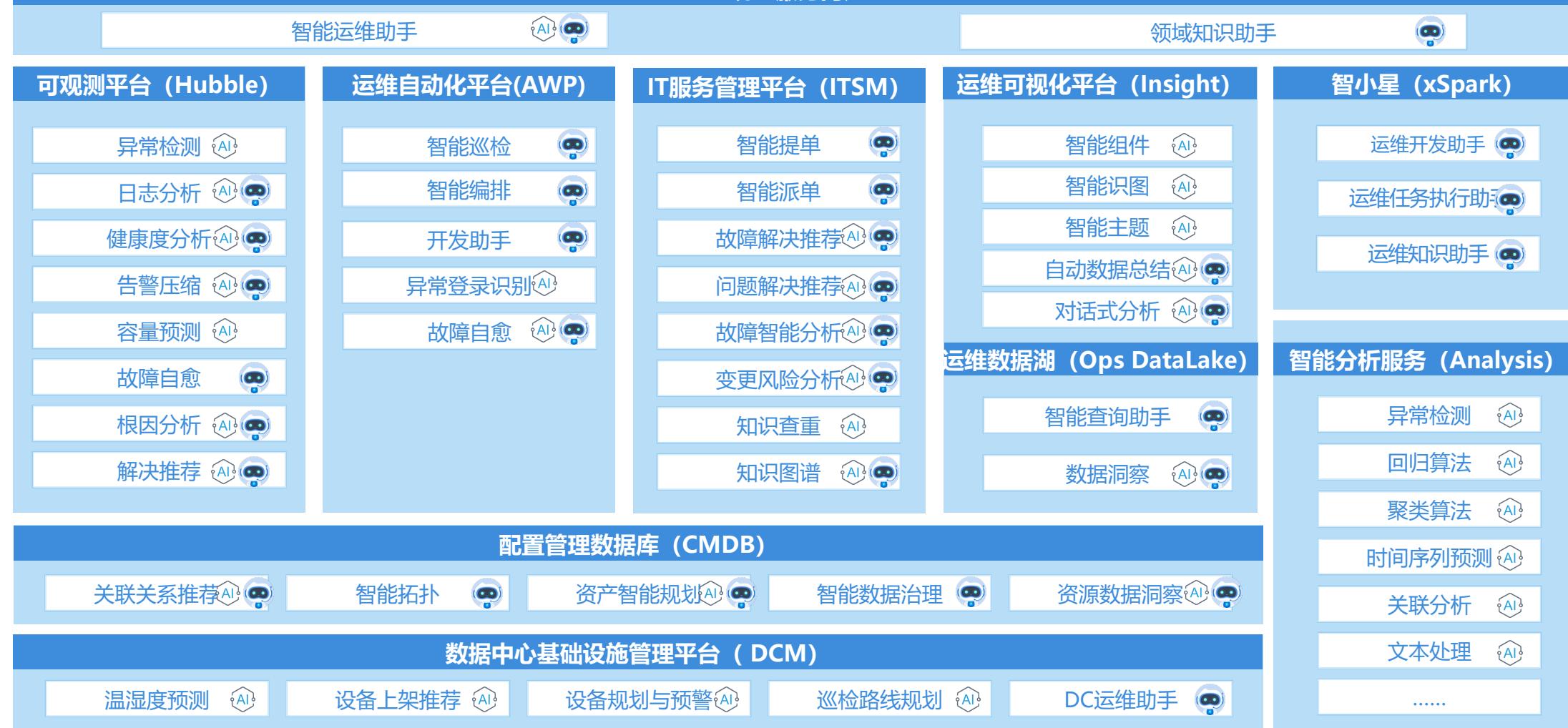
联想 xCloud AIOps - 智能运维场景总览

传统人工智能

生成式人工智能

联想 xCloud AIOps秉承 “AI as a Service, AI Everywhere” 的战略，全面整合判别式AI和生成式AI算法，提升智能运维能力。

统一服务门户



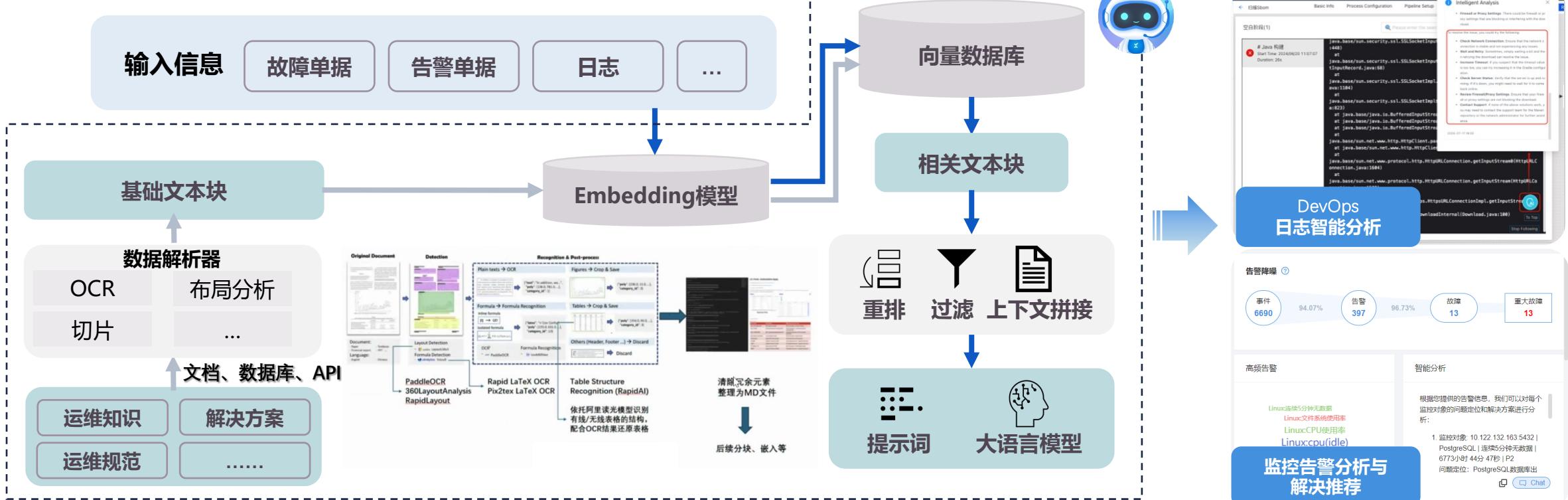
▶ 运维故障解决推荐

痛点及挑战

- 传统的事件诊断通常严重依赖专家领域知识。
- 通常需要从各种来源提取不同的数据，导致在识别和解决问题时出现延迟，影响系统性能和用户体验。

方案描述与价值

- 基于运维大模型，结合RAG技术，结合运维告警信息以及企业内部历史解决方案沉淀，提供定制化的洞察和准确的解决方案建议。
- 在警报和事件分析中提高了18%的效率，导致响应时间更快，整体系统可靠性得到提升。



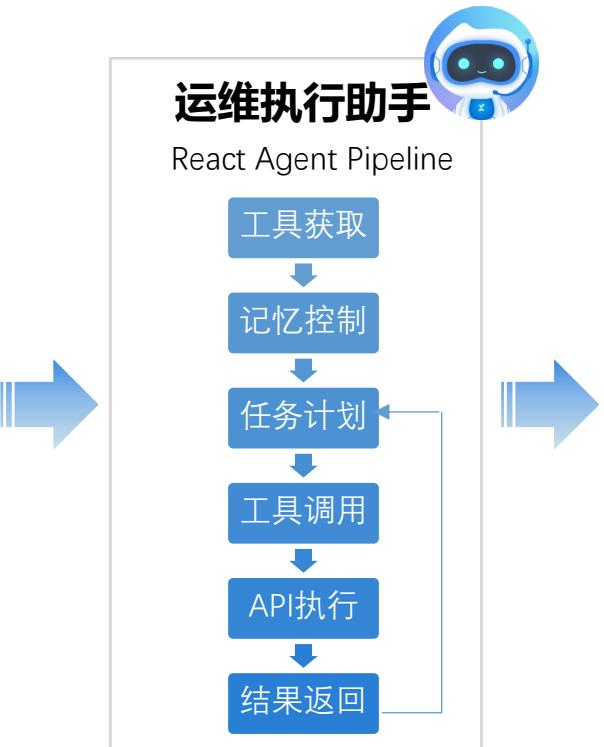
► GenAI IT自服务 - 智能运维执行助手

痛点及挑战

- 在IT运维中，存在大量重复性任务。
- 操作人员通常需要执行跨平台查询以满足一个简单的用户请求。

方案描述与价值

- 智能IT自服务：结合运维插件，用智能体实现用户自服务
- 提高用户自助服务率，减少操作人员的工作量。
- 在常规运维任务执行中提高了22%的效率。



查看智能运维大屏
2024-07-30 14:30
正在调用工具: Insight智慧大屏

运维报表查询
2024-07-30 14:30
正在调用工具: Service Control Platform

智能巡检
请利用AWP平台, 对服务器10.10.10.10进行巡检
2024-06-08 00:21
正在调用工具: Linux服务器巡检
您所提供的IP地址机器暂未收录在系统中。
2024-06-08 00:21

请利用AWP平台, 对服务器10.122.129.228进行巡检
2024-06-08 00:21
正在调用工具: Linux服务器巡检
您的报告生成成功! 点击查看: <https://xcloud.lenovo.com/awp/pages/actions/result?executionId=12f943d117394c86b5d737f010d609a>
2024-06-08 00:21

单据自助查询
2024-07-29 17:45
正在调用工具: 服务控制台请求查询

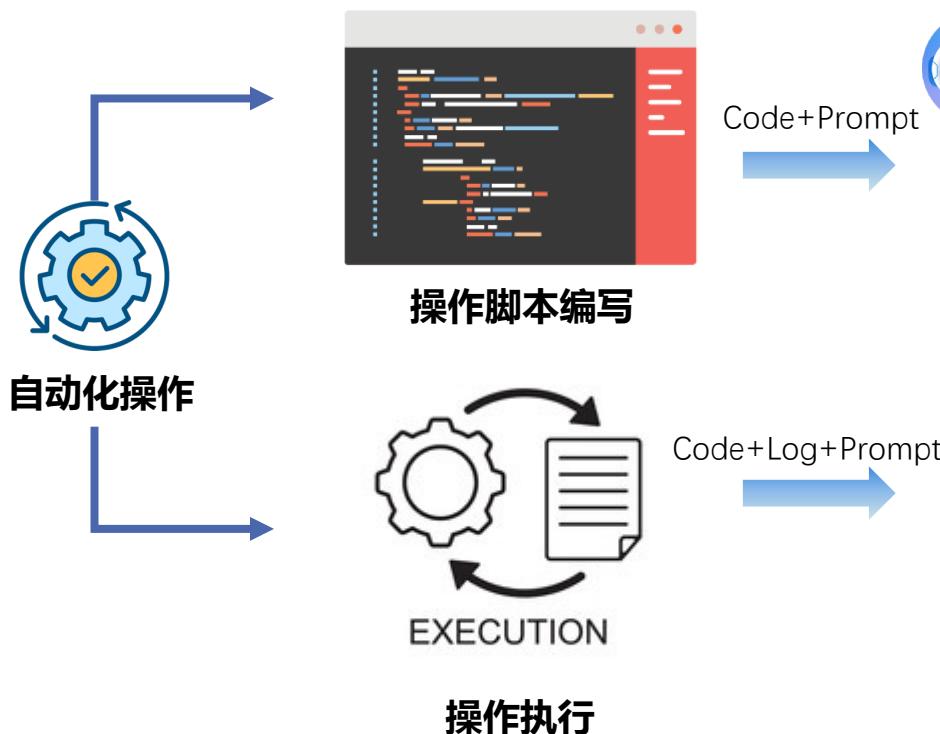
单号	服务名称	服务类别
CONCLOUDPU20240711181031763396198550	Container Project Expansion 容器云扩容	Container Service 容器云服务
CONCLOUDPU20240711173741005885198528	Container Project Expansion 容器云扩容	Container Service 容器云服务
FPS2024071113612876285198392	Firewall Policy 防火墙策略	NetWork 网络
FPS2024051019533202168187767	Firewall Policy 防火墙策略	NetWork 网络
VM220240205180830819873173133	Lenovo Cloud Server 联想云服务器	Computing 计算服务
CONCLOUD20231103162702644090159014	Container Cloud 容器云	Container Service 容器云服务
AC2023091211149283195150964	Application Certification 应用认证	Application Service 应用服务
VM220240205110406211107149753	Lenovo Cloud Server 联想云服务器	Computing 计算服务
NASC20230530095542211348133290	Create NAS Storage NAS存储创建申请	Storage Service 存储服务
VM22021102914562405808462226	Lenovo Cloud Server 联想云服务器	Computing 计算服务

巡检报告
下载巡检报告
巡检操作: 服务器-巡检-Linux
开始时间: 2023-10-26 17:52:49
结束时间: 2023-10-26 17:52:49
巡检涉及 1 个IT对象, 共计 1 个巡检实例, 识别风险项总如下:
巡检对象 实例类型 风险项 已开启巡检项 巡检结果 巡检状态
AWP_INSPECT_SERVER Linux 10.122.129.228 3 22 异常 SUCCEEDED

► GenAI智能自动化应用场景

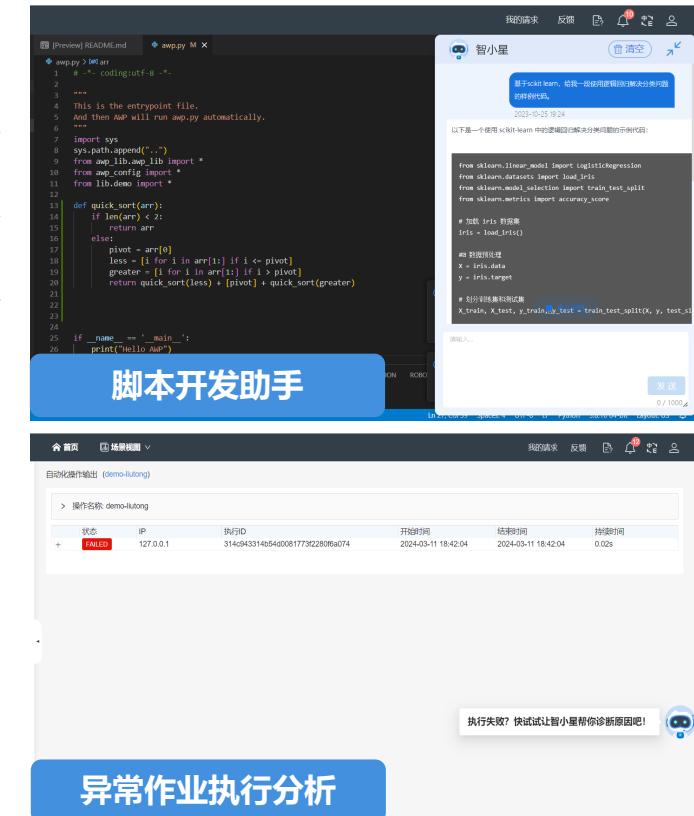
痛点及挑战

- IT运维经常需要创建和维护各种脚本，用于在不同系统上执行安装、部署和备份等任务。
- 手动开发脚本大量占用人力开发资源。



方案描述与价值

- GenAI支持的运维脚本开发，由专用代码模型驱动。
- 辅助生成自动化操作脚本：自动生成代码、智能建议和自动填写注释。
- 运维脚本开发效率提高了11%，降低运维编码门槛。



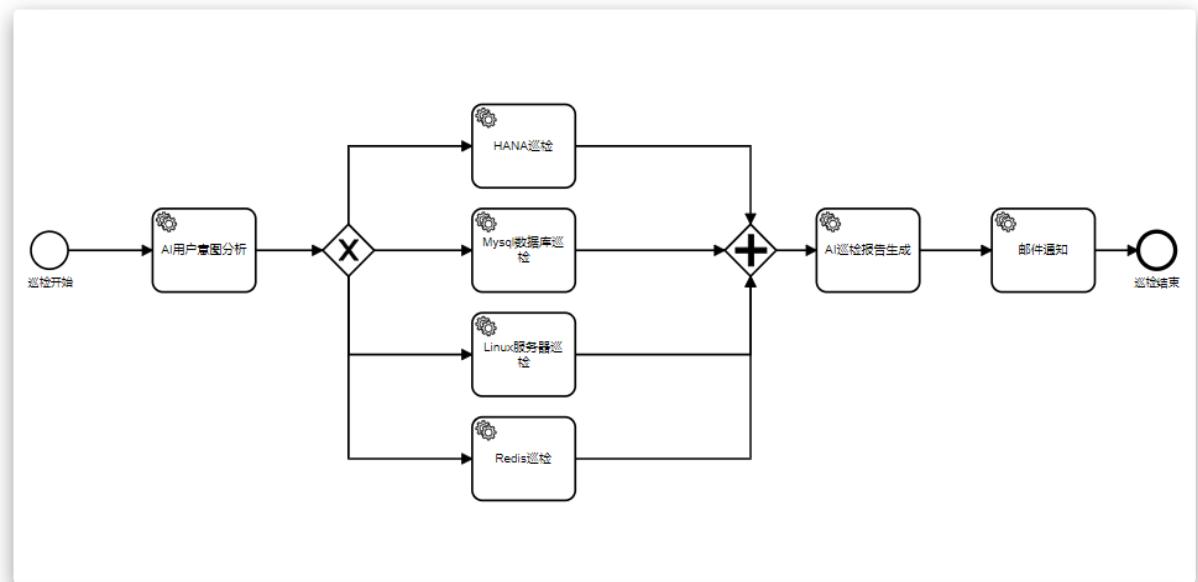
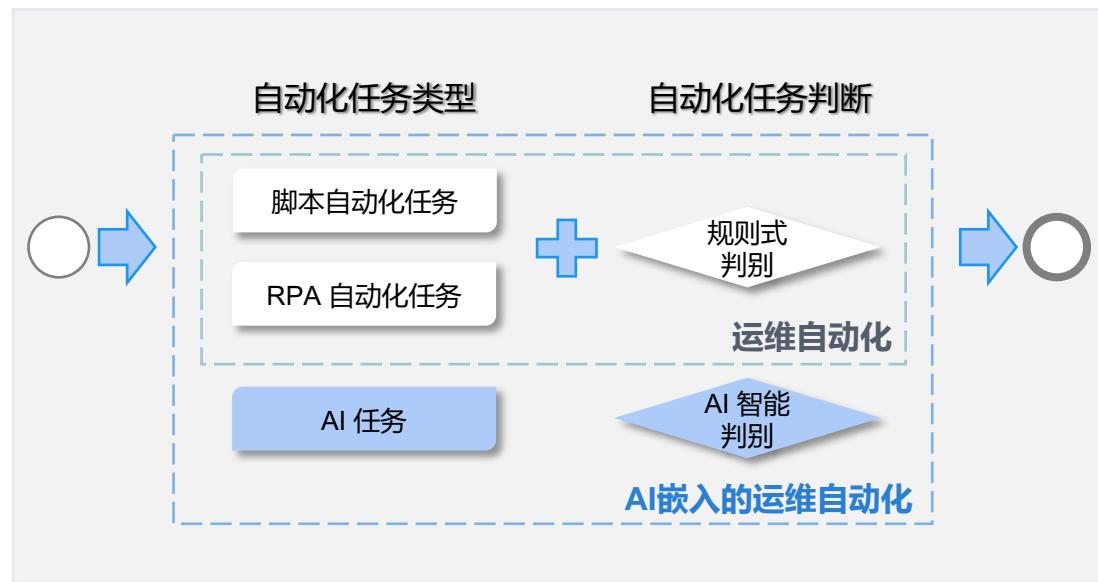
► GenAI智能自动化应用场景

痛点及挑战

- 数字经济带来员工与应用交互的场景数量上涨，对RPA技术需求在不断增长。
- RPA低代码配置技术比较僵硬，需要编码判断规则。

方案描述与价值

- 辅助RPA脚本的自动生成，同时，模拟人类大脑赋予认知能力，在RPA中起到分析与决策作用，实现更加灵活的智能自动化。



AI嵌入的
运维自动化场景

用户输入的意图分析分类

巡检报告总结

自动化任务告警内容生成

运维数据分析总结

...

► GenAI智能可观测性应用场景

痛点及挑战

- 多样化且分散的监控工具，缺乏端到端的监控视图和自助监控。
- 传统监控运维，需要大量依赖规则和专家经验，故障排查困难。

方案描述与价值

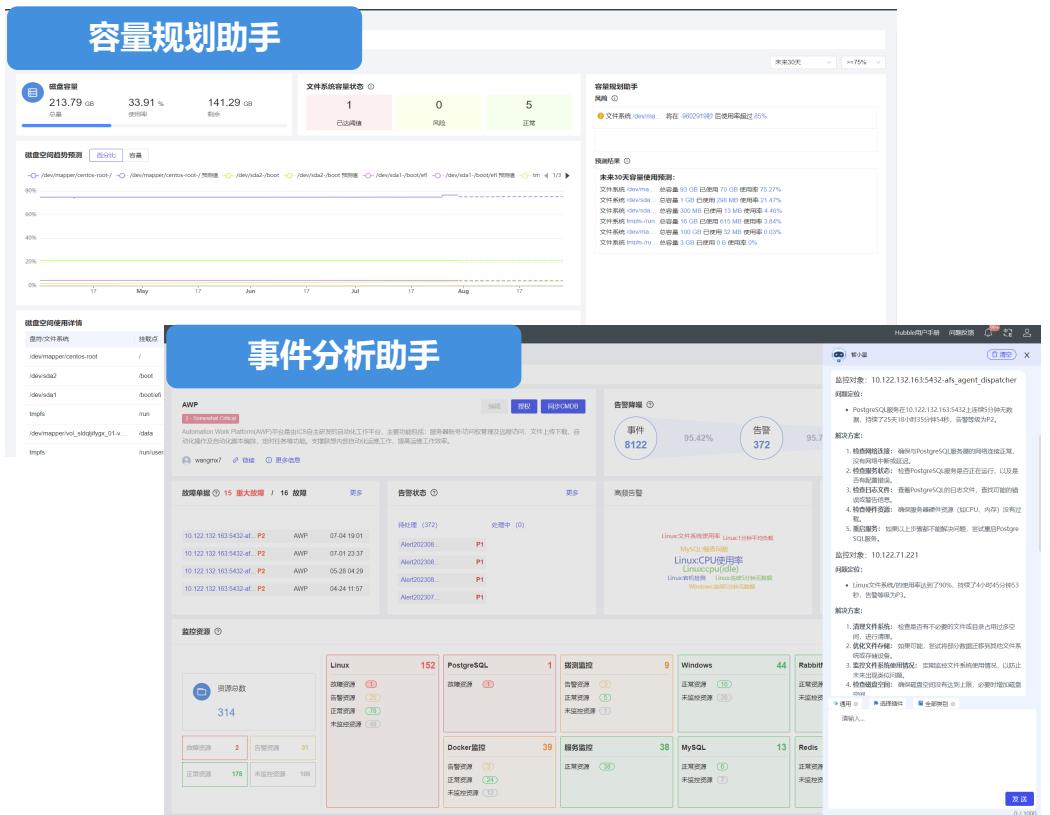
- AI赋能可观测性的全生命周期，提升系统内部的状态、行为和性能等方面观察、分析和监控的程度。
- 利用混合AI策略，涵盖事件事前、事中和事后不同阶段，提升运维处置效率。



事前：健康度评估、指标趋势预测等辅助IT运营风险防控、容量预测辅助资源精准规划、IT运营成本控制，降低故障发生数量；

事中：告警智能收敛及解决方案智能推荐等提升IT资源可观测性及告警效率，运维人员实时响应和处理，减少故障持续时间；

事后：根因分析、故障定位、影响面分析等辅助运维人员深入了解问题根因，提升运维质量及效率。



► GenAI运营数据洞察应用场景

痛点及挑战

- 大量的数据分散在数据孤岛中，需要快速进行数据挖掘与展示，发现数据的价值以指导运营。

数据ETL

Analytics Developer

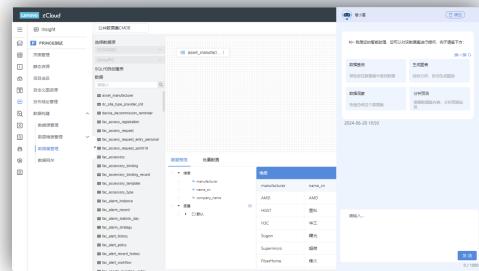
SQL脚本开发



Business Consumer

数据查询助手

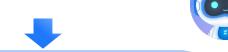
通过拖拉拽配置方式，即可完成数据接入与建模。配置数据查询助手，自动生成查询SQL。



数据探索性分析

Analytics Developer

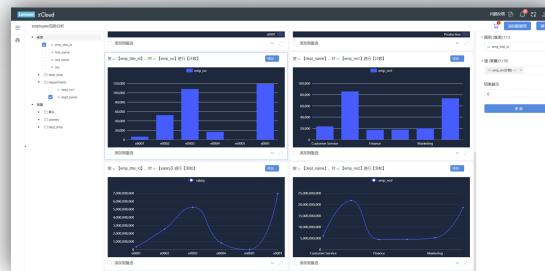
Python脚本开发



Business Consumer

组件智能推荐

仅需勾选字段，AI引擎自动推荐字段组合和聚合方式，并用合适的图表进行展示，快速构建图表。



方案描述与价值

- 利用生成式AI技术，打造运营数据查询与分析助手，让用户轻松实现运营数据报表构建与分析洞察。

报表配置

Analytics Developer

大屏&数据配置



Business Consumer

一键生成 & 智能美化

推荐图表加入备选后，可一键生成大屏。借助智能主题可进行配色、布局、边框元素的优化。



数据分析

Business Consumer

人工分析数据趋势



Business Consumer

对话式智能分析

基于决策式AI，对数据进行趋势、对比分析等。利用生成式AI，对话式交互，生成分析报告。



PART 04

实践经验与展望

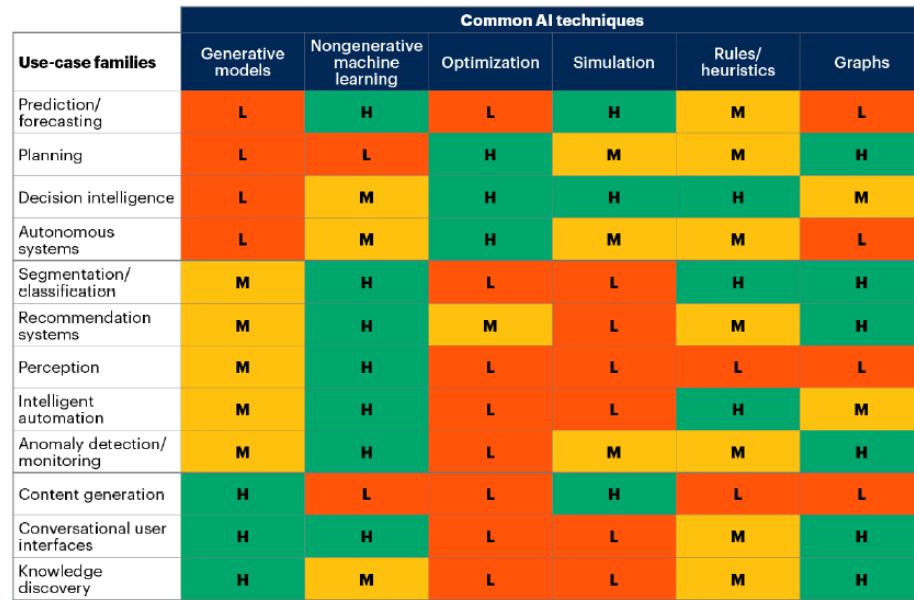
► GenAI智能运维应用实践经验 – 场景与模型选择

总结

- 结合GenAI擅长的领域(内容生成/总结，RAG，Code生成等)以及运维实际场景，选择采用合适的GenAI方法。
- 开闭源模型各有所长，不同场景选择不同模型。
- 数据安全会影响模型的选择(开源/闭源，私有化部署/公有云)。
- 构建模型运维追踪体系，通过量化指标定义模型表现，持续优化提升AI场景表现。

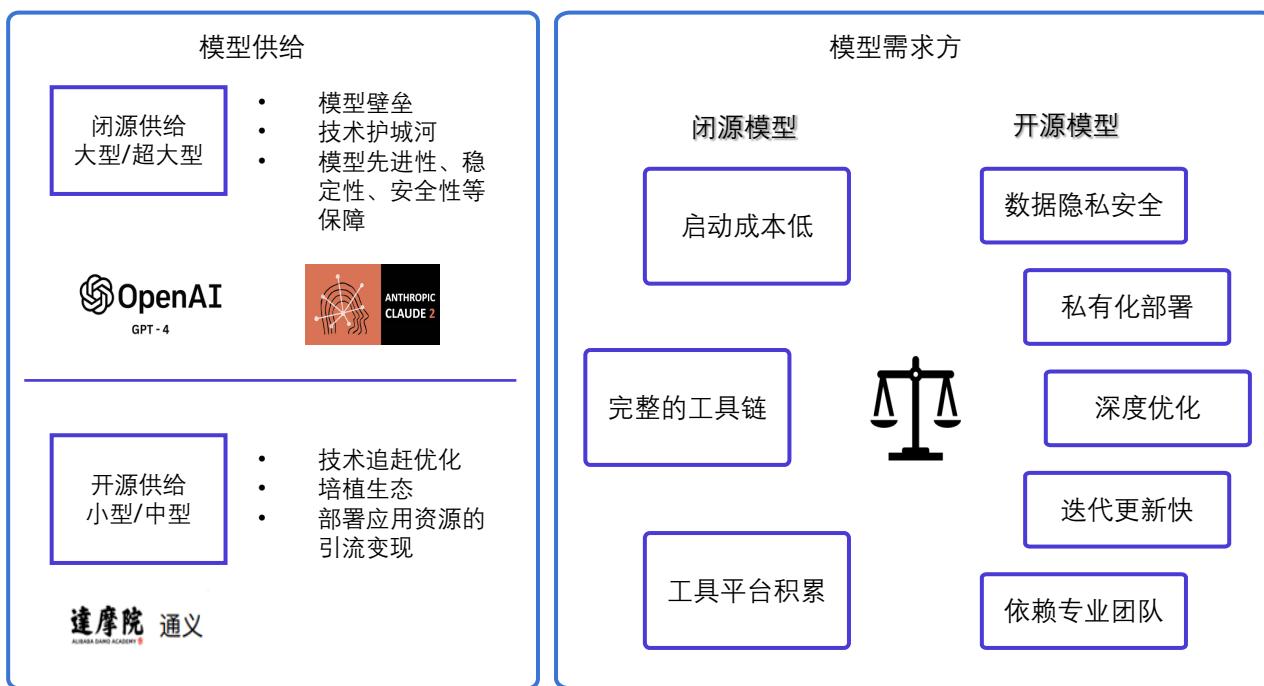
AI Techniques Heat Map

AI technique suitability ■ Low (L) ■ Medium (M) ■ High (H)



Source: Gartner
806238_C

Gartner



► GenAI智能运维应用实践经验 – 工程落地

总结

- AI Everywhere: GenAI嵌入到产品的实际功能场景中，触手可得，发挥价值；
- 持续运营，效果监测，不断优化（提示词/知识库/Fine Tune等）。
- 企业为确保生成式AI顺利落地，应协同产品经理和研发团队，推动一体化：包括体验一体化、数据一体化、架构一体化、流程一体化这四个主要方向；
- 提供统一入口、可扩展、可复用、可组装的一体化架构。



体验一体化

提供一致的连续的用户界面使用体验

数据一体化

统一并建立核心数据/权限模型，实现平台数据打通，保证数据一致性

架构一体化

设计一体化的技术架构，实现可复用的可组装的平台技术架构

流程一体化

针对用户端到端场景，设计连续的业务流程，打通各组件模块，实现一体化的流程

RAG在AIOps领域遇到的挑战与解决方案

挑战

问题描述

解决方案

其他手段



模型的过度生成

- 大多数内部用户的查询都是直接且精确的，这需要基于搜索的答案，而不是由LLM进行复杂的创造。

- 优先进行RAG入口的意图识别；
- 利用基于场景的路由，匹配知识库；
- 使用单一任务学习模型（SLMs）实现准确响应。

- 提示词工程
- 事实审查机制

术语理解差

- 无法正确理解企业内部的专用术语或者运维行业术语，细分行业知识。

- 开发一个针对运维专业术语和缩写的领域词汇表，作为全局知识，以提高模型在输入和检索中的理解能力。

- 模型微调
- 工具调用
- 知识图谱

无法提供用户满意答复

- 尽管检索了相关的数据，RAG的生成可能在提供全面、精确和有益的答案方面仍有不足。

- 扩充知识库，加强与已有知识资源的集成；
- 提升模型能力，减少模型制约：如超大文档；
- 提升用户查询方法。

- 构建用户反馈机制

企业的敏感话题

- 对公司品牌、形象和高管话题的不准确回应可能导致危机。

- 在公司层面建立一个包含标准化回答的敏感话题资料库。

- 知识管理
- 模型微调
- 增加LLM输入、输出护栏

信息呈现与理解

- 图表和附图解读难、分段信息乱

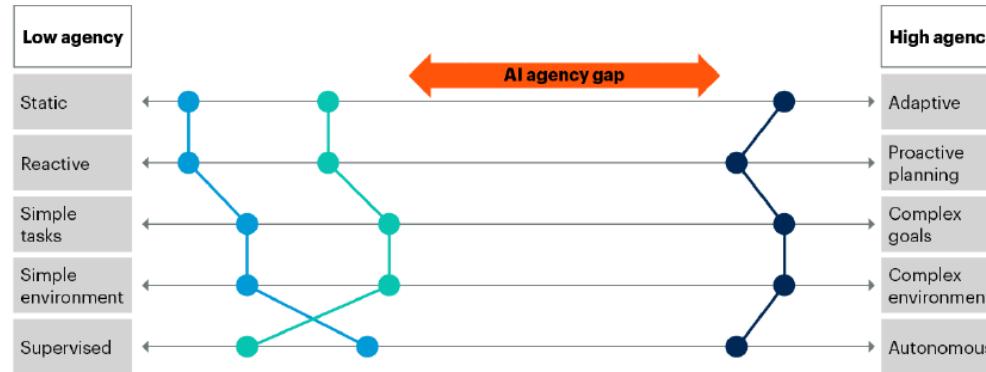
- 通过大模型和OCR，提取文档附图和表格的关键信息；
- 配合文档元数据解析，配合专用的布局分析模型，实现文档结构的准确理解。

- 多模态大模型

▶ 智能体在AIOps领域遇到的挑战与解决方案

Mind the AI Agency Gap

● Human agency ● Deterministic chatbots ● LLM-based assistants



Source: Gartner
806843_C



► GenAI智能运维应用展望

随着GenAI的发展，应用场景将逐渐从嵌入模式，助理模式升级到智能体模式，IT运营智能体将助力企业实现智能卓越运营





THANKS

