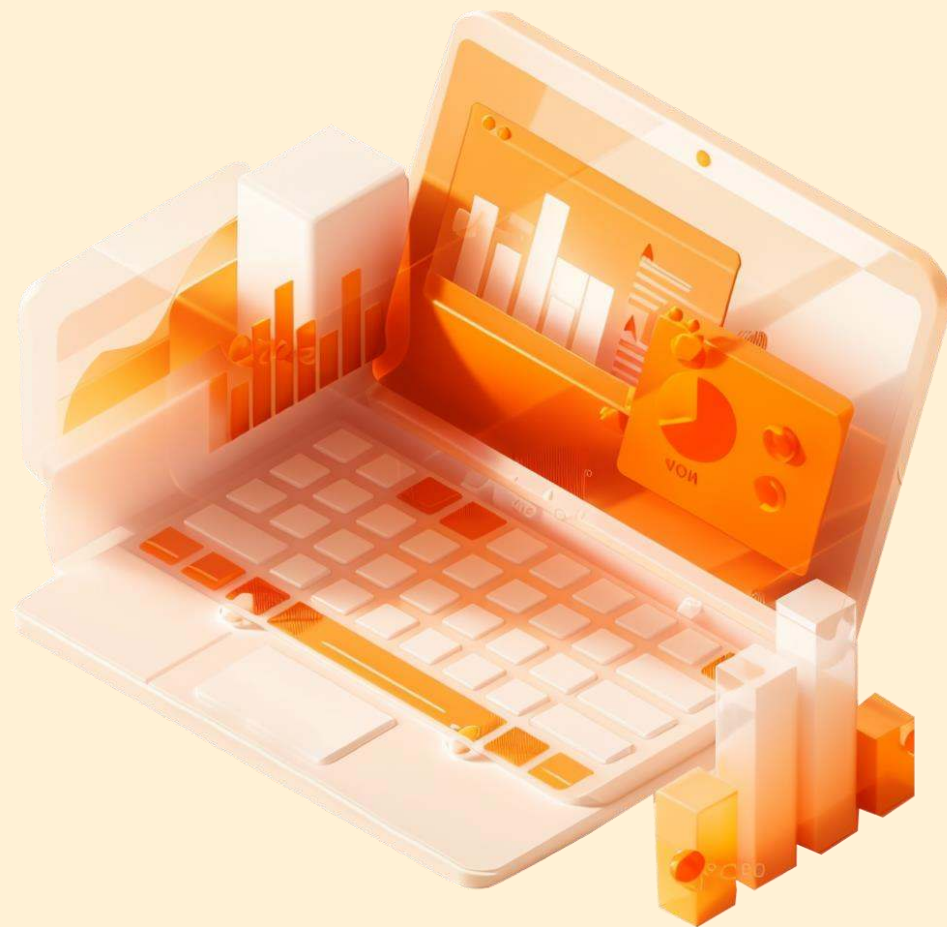


AI助力炎凰腾飞

新一代多源异构大数据数智化平台



- 公司与产品简介
- 数智化背景：AI创新焦虑
- 炎凰数智工具箱 
- 炎凰数智工具箱实践
- Q&A

上海炎凰数据科技有限公司



炎凰数据作为具有国内完全自主日志分析产品和先进技术路线的开创者和引领者，拥有一支由前 Splunk 中国研发中心核心团队成员组成的优秀团队。



炎凰数据**突破国外核心技术壁垒，填补了我国日志分析产品独立自主可持续发展技术路线的空白，摆脱了国外核心技术路线的依赖。**

01 公司团队



炎凰数据创始团队来自 Splunk

董事长 何宁 (中间) - 原 Splunk 上海研发中心总经理
首席运营官 石俊 (左一) - 原 Splunk 上海高级软件开发经理
首席架构师 倪跃 (左二) - 原 Splunk 上海首席软件工程师
首席技术官 尤天 (右二) - 原 Splunk 上海首席软件工程师
研发副总裁 路铭昊 (右一) - 原 Splunk 上海高级研发经理



整个团队建制完整，核心技术团队磨合时间超过10年，且具有丰富和深入的非结构化数据库以及即时数据分析引擎内核开发经验。

01 公司团队



70%

拥有硕士及以上学历



80%

来自国内外顶尖院校



90%

技术研发人员占比



10年+

核心团队建制来自前
Splunk中国研发中心，
具有深厚大数据分析、架构
设计和系统开发经验

上海炎凰数据科技有限公司作为具有完全自主日志分析产品和先进技术路线的开创者和引领者，拥有一支由前Splunk中国研发中心核心团队成员组成的优秀团队。我们突破国外核心技术壁垒，填补了我国日志分析产品独立自主可持续发展技术路径上的空白。

数智时代背景下，炎凰数据致力于使“每个人轻松快速发现并发挥数据要素价值、创造价值场景”。通过炎凰数据分析平台一站式完成数据的采集、存储、分析、展示与共享，低代码快速实现各种数据价值应用场景落地，帮助企业用户成功实现数智化转型。炎凰数据业务场景广泛，包括但不限于：智能运维、信息安全、审计、业务分析、流程优化、用户行为分析、智慧城市、物联网服务等多个场景，覆盖金融、制造、能源、交通、医疗、教育等众多行业客户。

投资方

晨山資本
MOUNT MORNING CAPITAL

SEQUOIA 红杉

上海云基地
SHANGHAI CLOUD VALLEY

蓝驰
BLUERUN
VENTURES

信雅达
SUNYARD

朗玛峰创投

容亿投资
WINREAL INVESTMENT

新尚资本
SHANG CAPITAL
政府引导基金

01 客户与合作伙伴



服务客户

银行



保险



证券



能源/制造



其他行业



炎凰数据驱动业务价值-让数据应用更简单

应用与价值

物联网

- ✓ 运管
- ✓ 检测

开发运维

- ✓ 效能
- ✓ 绩效

IT 运维

- ✓ 监控
- ✓ 告警
- ✓ 排障

安全运维

- ✓ SIEM
- ✓ 态势感知

用户行为分析

- ✓ 日志与埋点
- ✓ 数据互补

商业分析

- ✓ 营销
- ✓ 客服
- ✓ BI报表

流程挖掘

- ✓ 流程评估
- ✓ 流程优化

数据建模 与处理

优势一：Fast time to value

从数据接入到查询分析快速落地

优势二：强大的查询分析能力

Ad Hoc查询、即时查询、交互式查询、关联分析、自助式分析

优势三：SQL交互

标准SQL交互，算子强大（表函数、标量函数）

优势四：强大的扩展能力

用户可以自定义各种SQL表函数、Python表函数、SQL标量函数和Rust标量函数等

数据源

优势五：不挑数据源。不需要数据清洗，用户无需关心繁重的数据清洗和ETL工作

只要是文本数据，无论是结构化、半结构化、非结构化、混合结构等多源异构数据



一句话：一站式完成任何文本时序数据采集、存储、关联分析、展示与数据共享，低代码快速实现各种数据分析场景落地；让每个人都可以轻松创造数据价值！

核心优势

完整保留原始信息，实现一份数据的多视角、多维度、全方位观测与场景应用

统一开放平台，多种场景

读时建模Schemaless SQL 无与伦比的灵活性

读时建模，完整保留事实和细节，高效响应业务的同时，显著提升分析的准确性、有效性与合规要求

快速挖掘数据价值

数据采集、存储、分析、可视化与分享一站式快速部署，聚焦数据价值挖掘，快速提炼和展现数据要素价值



高效关联分析，实时洞察数据价值

动态引入内外部关联数据，扩展富化增强分析维度，增强推理分析和决策的全面性和准确性。同时，基于AI实现生成式富化查询。

可持续发展的独立自主技术路线

打破依赖国外核心技术依赖，拥有**国内唯一不依赖于**开源产品，完全独立自主的可持续发展基于云原生的日志平台技术路线。保持先进性、扩展性的同时，满足当下与未来技术架构发展需要。



快速挖掘数据价值

常规式数据分析时代

探索式数据分析时代

处理 数据 范围	5-10%	结构化	✓	✓
	5-10%	半结构化		✓
	80-90%	非结构化		✓

分析
数据
流程



沟通链条长

实施流程复杂

数据使用效率低

周期较长，以月为单位



通过新一代数据
分析引擎加速数
据资产价值实现

原始数据

周期短，以小时为单位

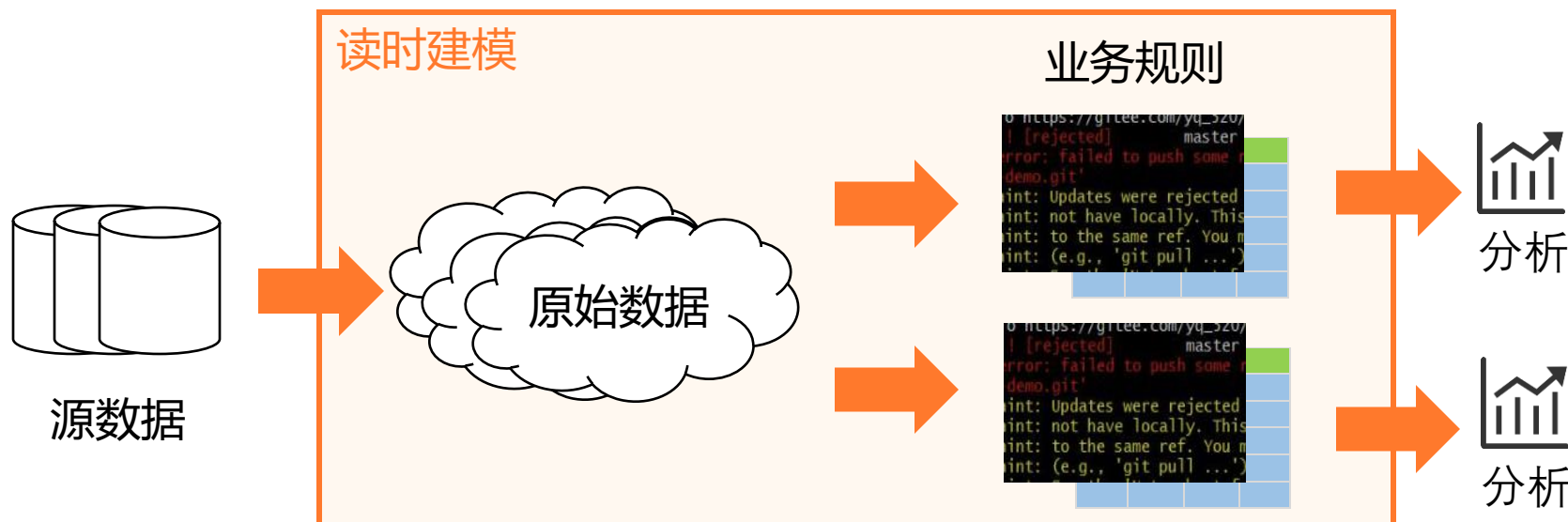
部署
周期

读时建模 Schemaless SQL 无与伦比的灵活性



写时模型，典型应用在传统数据库，数据在入库的时候需要预先设置schema，简单讲就是表结构

炎凰数据：读时建模+写时建模



读时模型，在数据汇聚存储到数据分析之间，数据先存储，然后在需要分析的时候再为数据设置 schema

混合建模



机器数据

Logs

Metrics

Traces

运营数据

基础数据

业务数据

行为数据

结构判断

半结构化数据

结构化数据

原始数据存储

数据索引

数据清洗

预定义解析

数据存储

动态解析

关联分析

字段读取

计算聚合

探索型即席分析

灵活性：业务需求多变

报表型常规分析

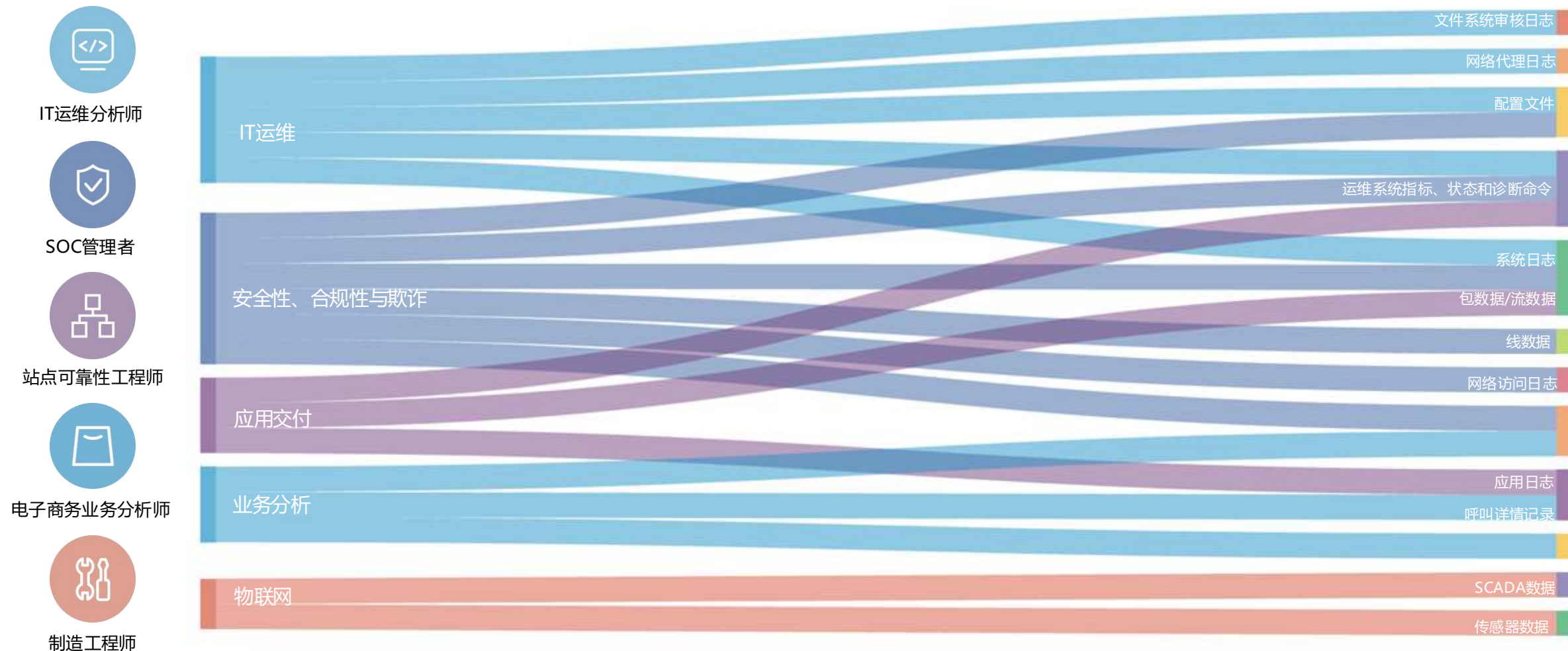
时效性：业务需求固化

读时建模
国内第一真正技术实现



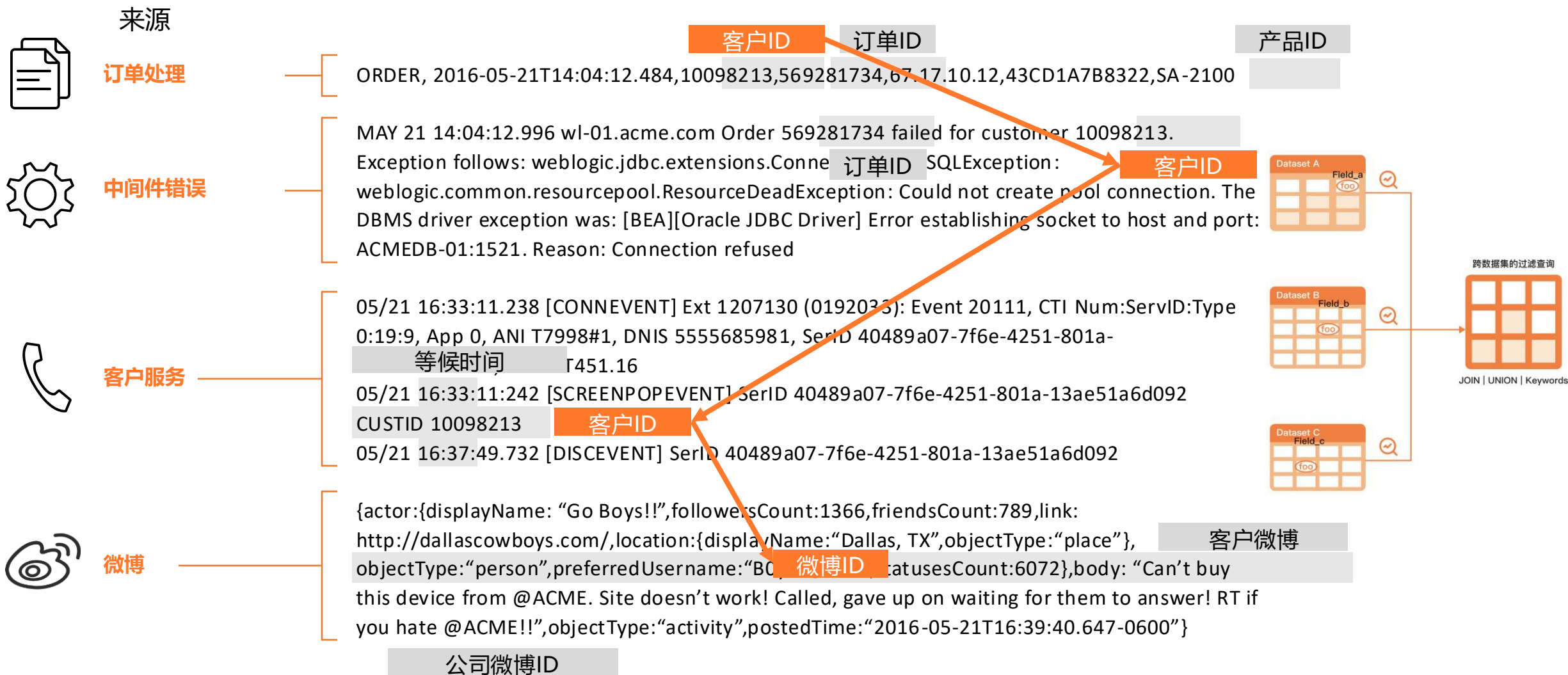
写时建模

01 统一开放平台，多种场景



通过 Schemaless SQL 实现一份数据可以应用在不同场景，而不需要额外增加数据冗余

高效关联分析，实时洞察数据价值

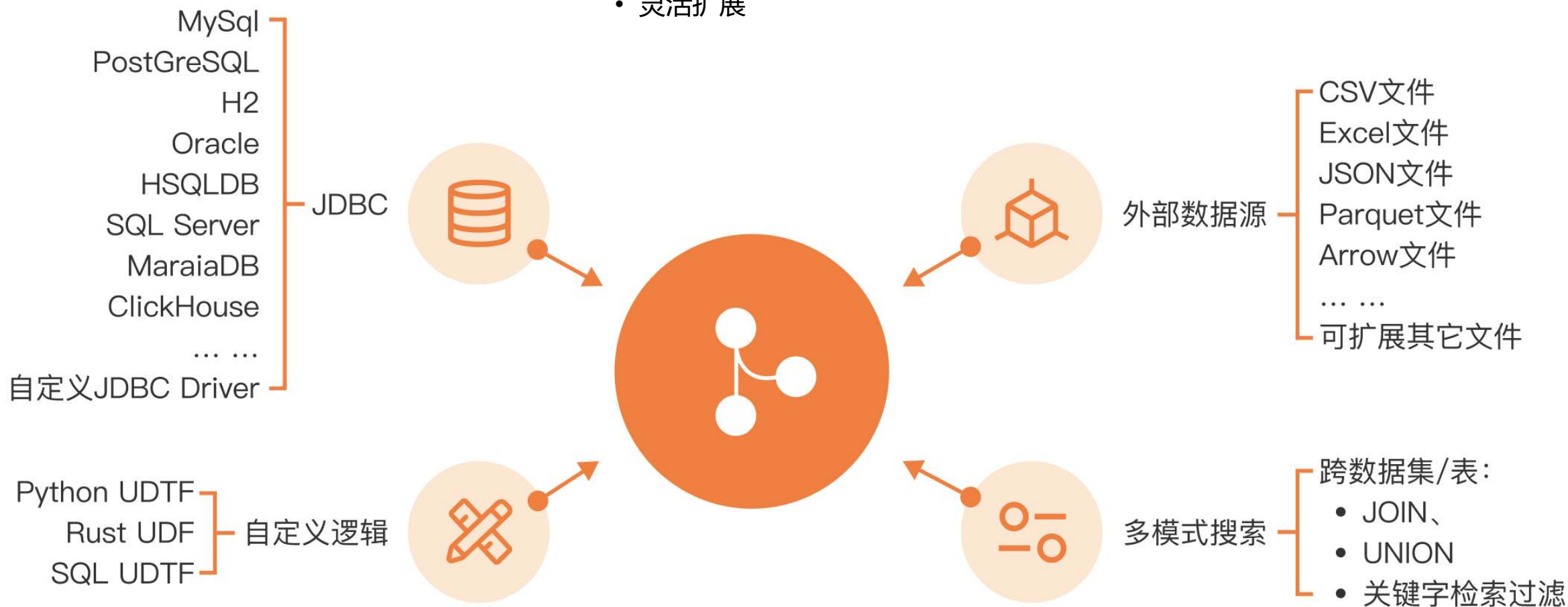


实现端到端，跨系统复杂关联分析与数据价值洞察

01 关联查询

优势

- 灵活关联外部数据源
- 无须了解外部数据表结构变化
- 内置SQL关联查询
- 灵活扩展



01 标准SQL



SQL分析引擎

采用标准SQL语法

- ✓ 无需学习专用的查询语言
- ✓ 兼容已有的SQL系统，减少二次开发的成本
- ✓ 生态成熟，更易于从互联网以及社区获得帮助



函数计算引擎

在SQL基础功能之上还提供丰富的高级分析功能，帮助用户应对复杂分析问题

- ✓ 标量函数(Scalar Function)
- ✓ 表函数(Table Function)
- ✓ 公用表表达式(CTE)
- ✓ 视图(View)
- ✓ 具备多数据源Join的功能，帮助用户进行多层关联分析
- ✓ 自定义扩展：SQL 标量函数、SQL 表函数、Rust标量函数以及Python表函数

```
1 WITH weibo_with_topic AS (  
2   SELECT weibo_demo.article, topic_table.topic  
3   FROM weibo_demo OUTER APPLY parse_regex(weibo_demo.article, '#(<?topic>.+)#') AS topic_table  
4   WHERE topic_table.topic IS NOT NULL  
5 )  
6 SELECT topic, count(*) AS message_count FROM weibo_with_topic GROUP BY topic ORDER BY message_count DESC
```



预计算加速器

利用物化视图加速表聚合、表连接

<支持标准SQL查询语言，基于Web UI快速扩展SQL UDTF/UDF，同时结合Rust和Python的UDTF/UDF扩展计算能力，通过视图和公共表表达式等抽象组合，精准表达查询意图>

可持续发展的独立自主技术路线





云原生微服务架构弹性扩展，低耦合，横向/纵向按需扩容



二次开发扩展

REST API接口以及二次开发
SDK扩展



采集器扩展

用户自定义DataScale采集器
插件与开源采集器自由扩展



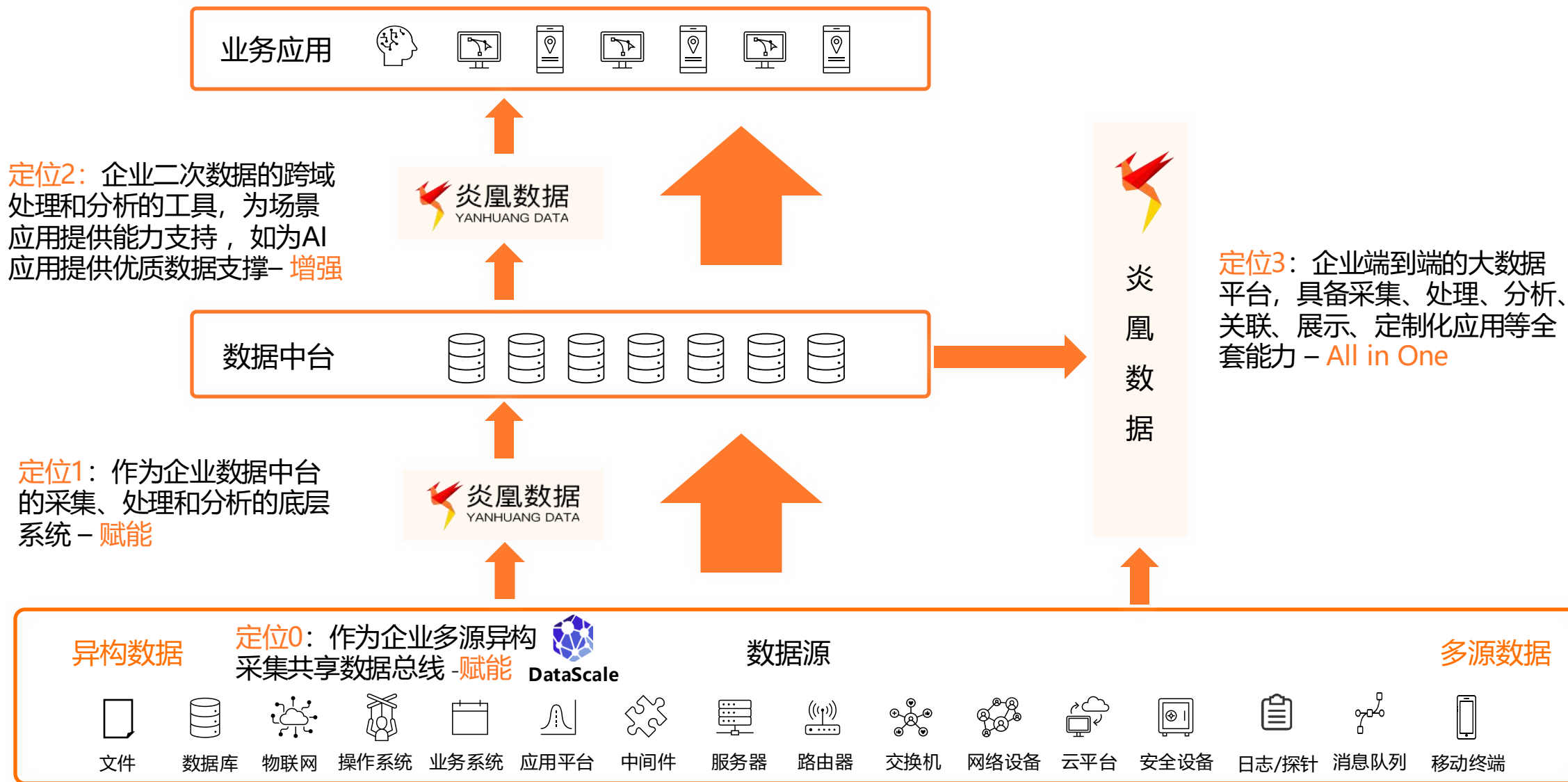
Web UI SQL表函数、
Python 表函数自由扩
展及Rust标量函数自
由扩展



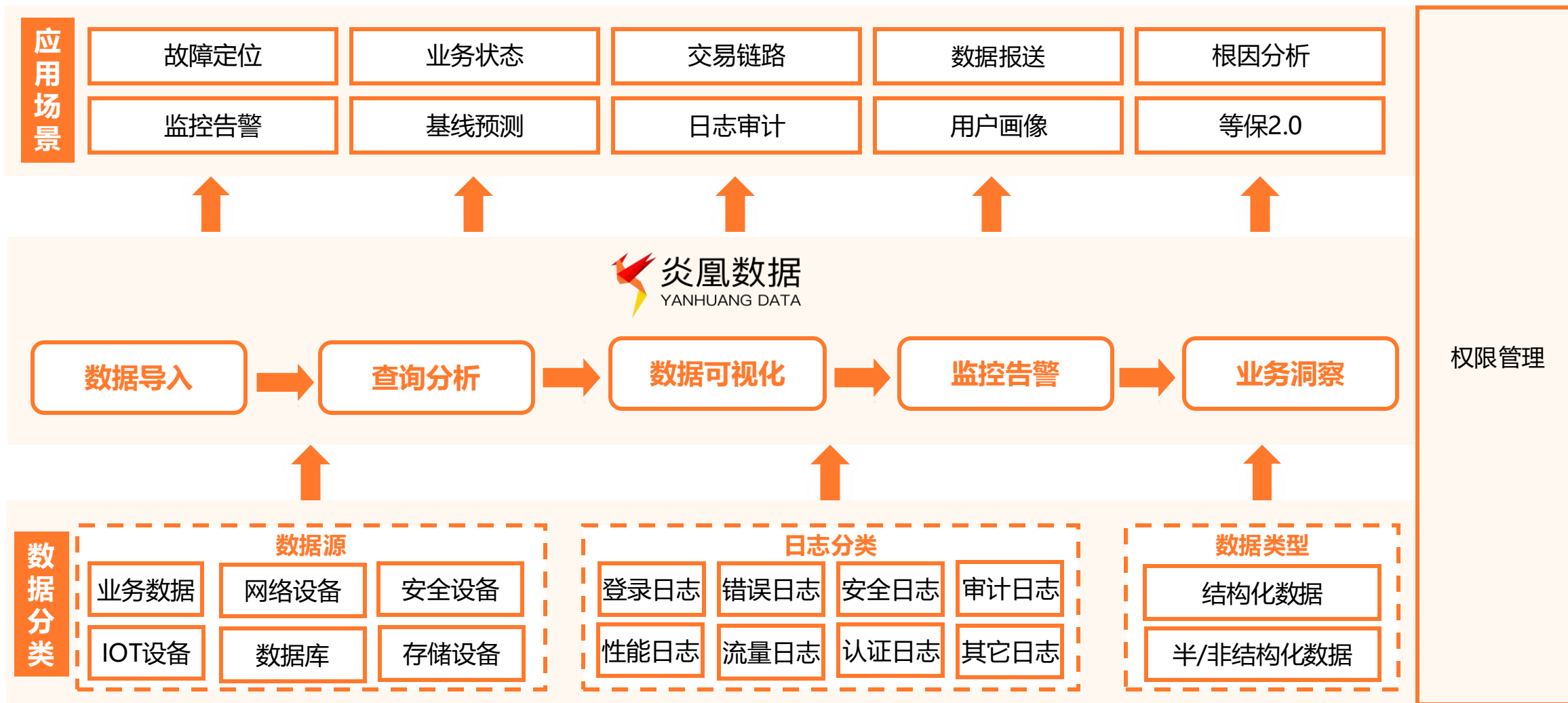
丰富扩展
函数扩展



产品定位



平台功能



用大数据分析的方法解决各种问题

一个平台，多个场景，多种价值

背景介绍

IT 运营

可观测性

安全与合规

业务分析

物联网

预测服务品质的下降，防患于未然

快速定位应用性能不佳的原因是由代码错误还是基础设施所导致

加快安全调查，减少内部威胁的影响

洞察营销活动是否通过网站或移动应用提升了订单数量

实时监控和分析数万个传感器的数据

02 背景：AI创新焦虑

- 近2年，大模型风生水起
- 如何在企业落地，我们一直在思考
- 客户对产品有了新的智能化要求
- 数字化向数智化转型
- 2025年1月DeepSeek更快更好的国产大模型为行业快速落地奠定坚实基础
- 全员“AI创新焦虑”



02 破解AI创新焦虑-3个假设

01. 把AI当"人"看

人工智能，关键在于“人工”，即希望达到人类的智慧水平，并为人服务。

02. 大模型/AI“幻觉”不可避免

正如人存在幻觉，误判等，基于人类神经网络发展而来的人工智能，必然也无法摆脱这一属性。正如现代企业中的工作流，SOP，都是为了实现标准化，高质量输出，避免人的因素导致质量下降。

03. AI不是替代人，而是用来辅助人

以目前AI发展的情况看，辅助人类将人从重复、低价值工作释放出来，进行更高价值的工作，是AI首要目标，并且会有更好的效果。



02 数智化创新（AI场景真正落地）关键



实现数智化创新落地的关键：

1. AI基础设施
2. 高质量数据基础
3. 快速搭建AI应用的低代码开发平台
4. 客户领域知识专家 (人才)：AI；数据；IT基础设施.....

02 价值场景判断标准

人效相关

是否可以用已有工具完成[如果存在，则是否定项]

当前现状是否必须人工参与，是否存在低价值、耗时费力的工作流程，可以优化[如果存在，是潜在的可改善的场景，作为肯定项]

成本可控

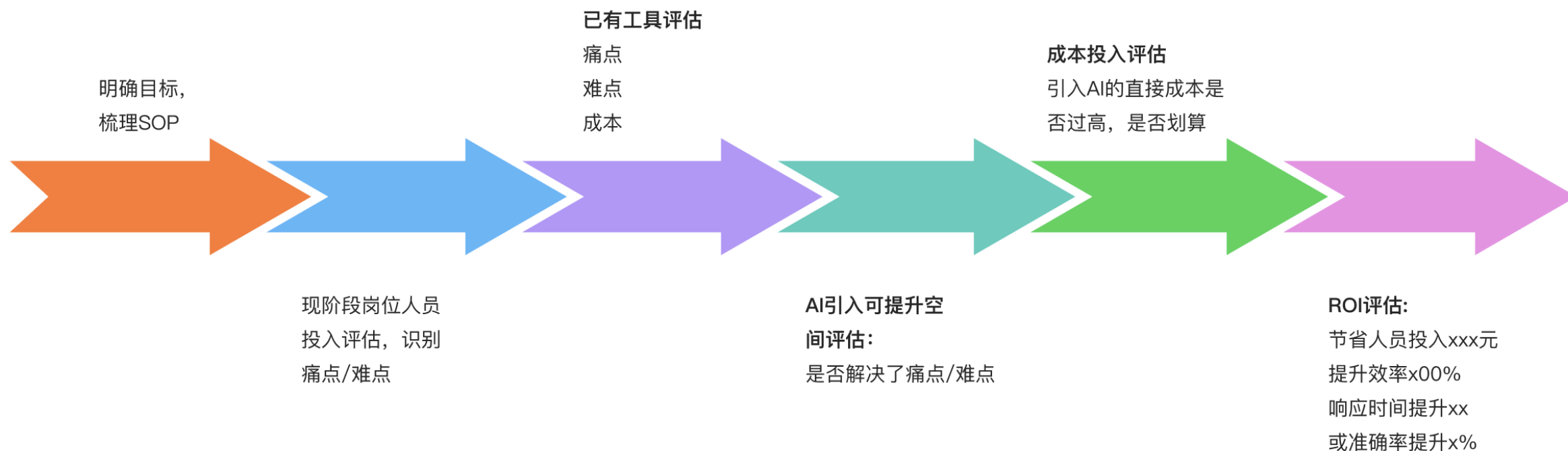
无底线的成本投入，不符合AI辅助人类降本增效/提质增效的目标[如果无法评估成本，则是否定项]

提高效能

确认场景落地后有明显的效能提升



02 SOP梳理：以人为核心，SOP为手段，难点/痛点为切入点

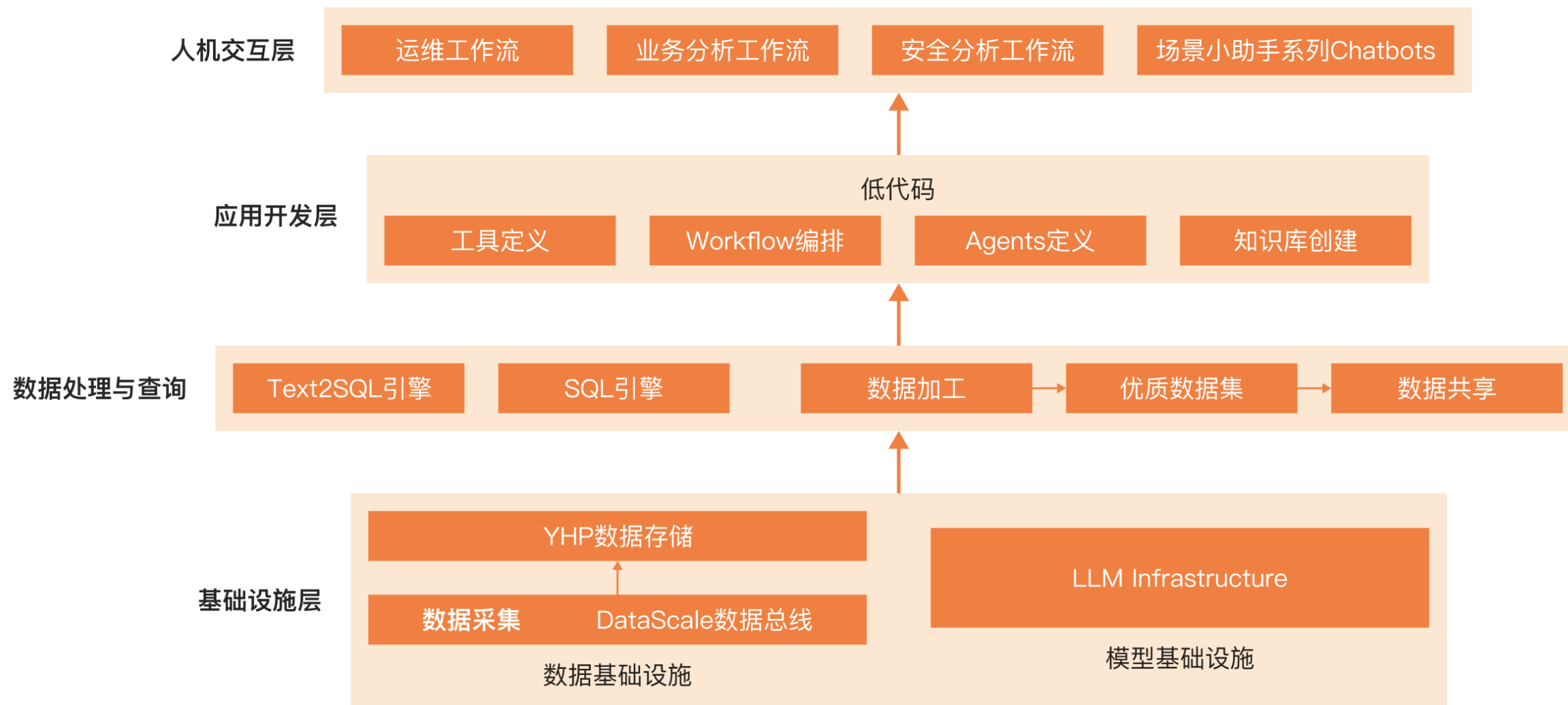


一个问题始终无法绕过：这个场景为什么要用AI？投入会获得预期的收益吗？
越早回答这些问题，去伪存真，更容易找到高价值场景。

03 炎凰数智工具箱 - 解决AI场景落地难的问题 帮助伙伴/客户，加速客户AI场景验证与落地



03 数智工具箱功能架构示意图



03

高质量数据基础：

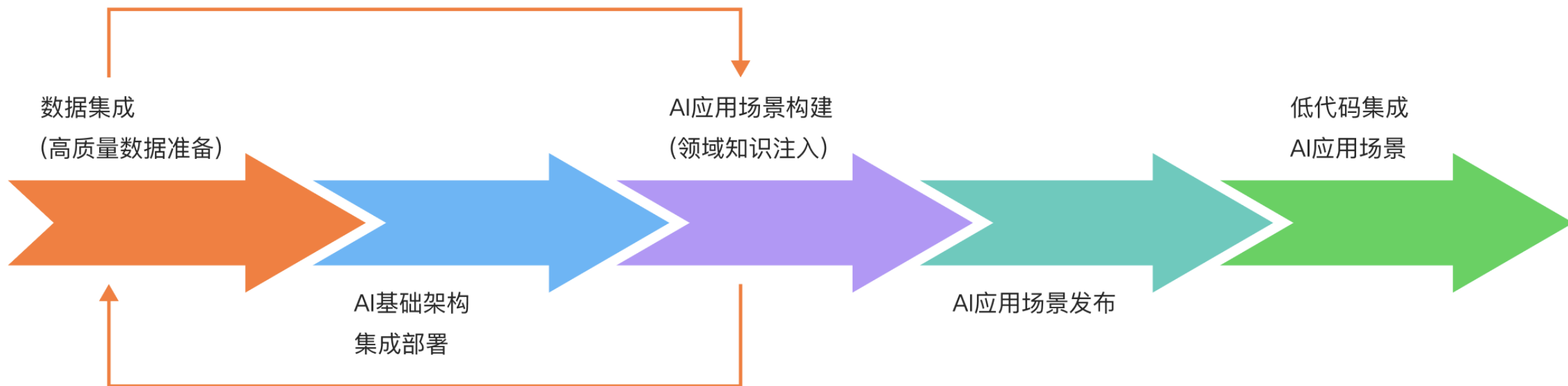
YHP端到端数据采集、存储加工、分析与共享



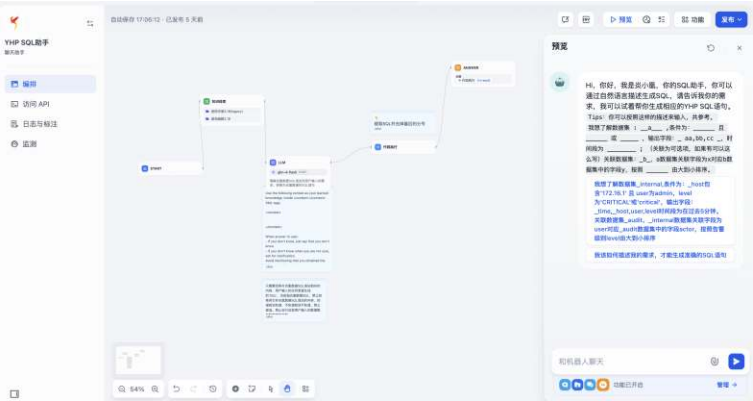
炎凰数据底座YHP解决传统数据处理与准备过程中的痛点：

- 传统的数据接入需要大量的ETL工作，这为本就繁重的数据准备环节增加了工作量；
- 灵活性不足：通过代码开发编写的方式，无法处理多源异构下，不同日志处理的需要；无法按需随时提取新的字段/标签/特征，无法快速根据不同模型提供针对性的高质量数据，开发周期长
- 标准化不足：缺少统一的数据管理，对于基础分析的日志观察缺少可视化分析工具辅助理解数据，并提取处理相应字段；
- 实时性不足：实时数据接入与存储，方便随时有基础数据用于各种新场景

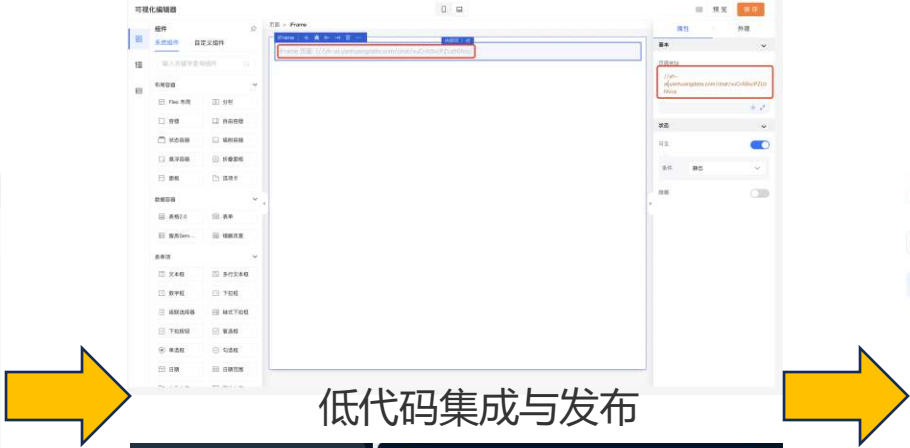
04 数智化场景落地实践



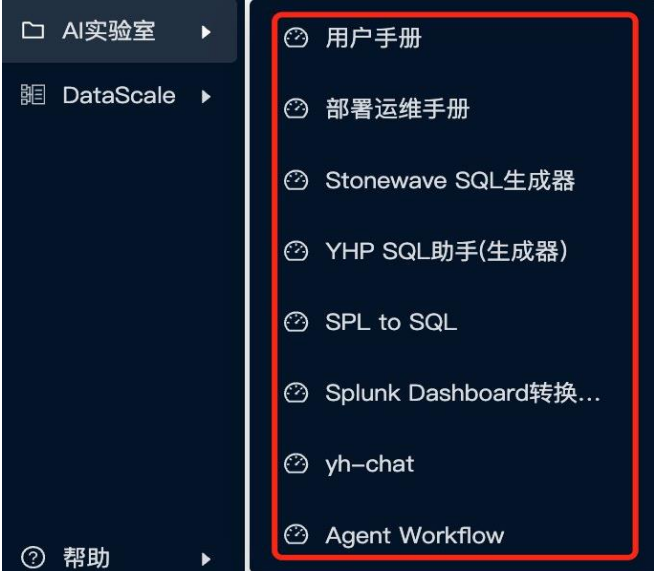
AI应用快速集成：1分钟完AI应用发布 实现数智场景快速落地



可视化工作流构建



低代码集成与发布



AI应用交付

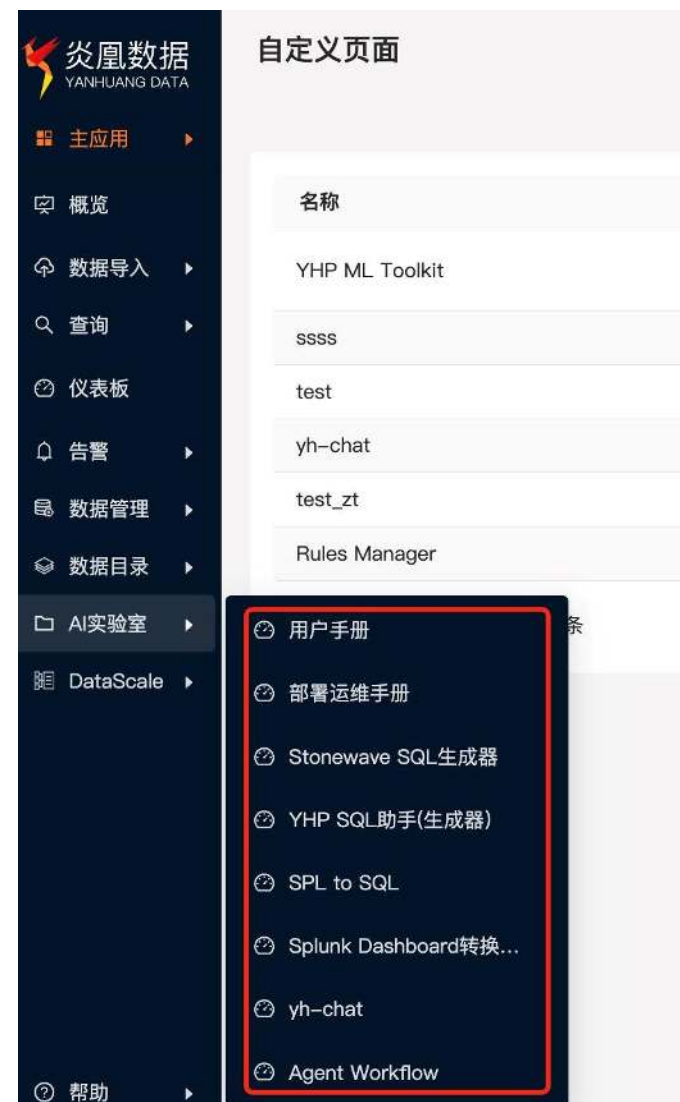
04 基于炎凰数据自己领域知识的探索-典型场景

场景落地

- 知识库
- Text2SQL
- SPL转换助手
- Splunk SimpleXML to YHP Dashboard JSON

YHP集成

- YHP低代码可视化编辑器-快速集成AI iframe
- YH dev (AI实验室)



04 典型场景一 炎凰产品使用和运维知识库

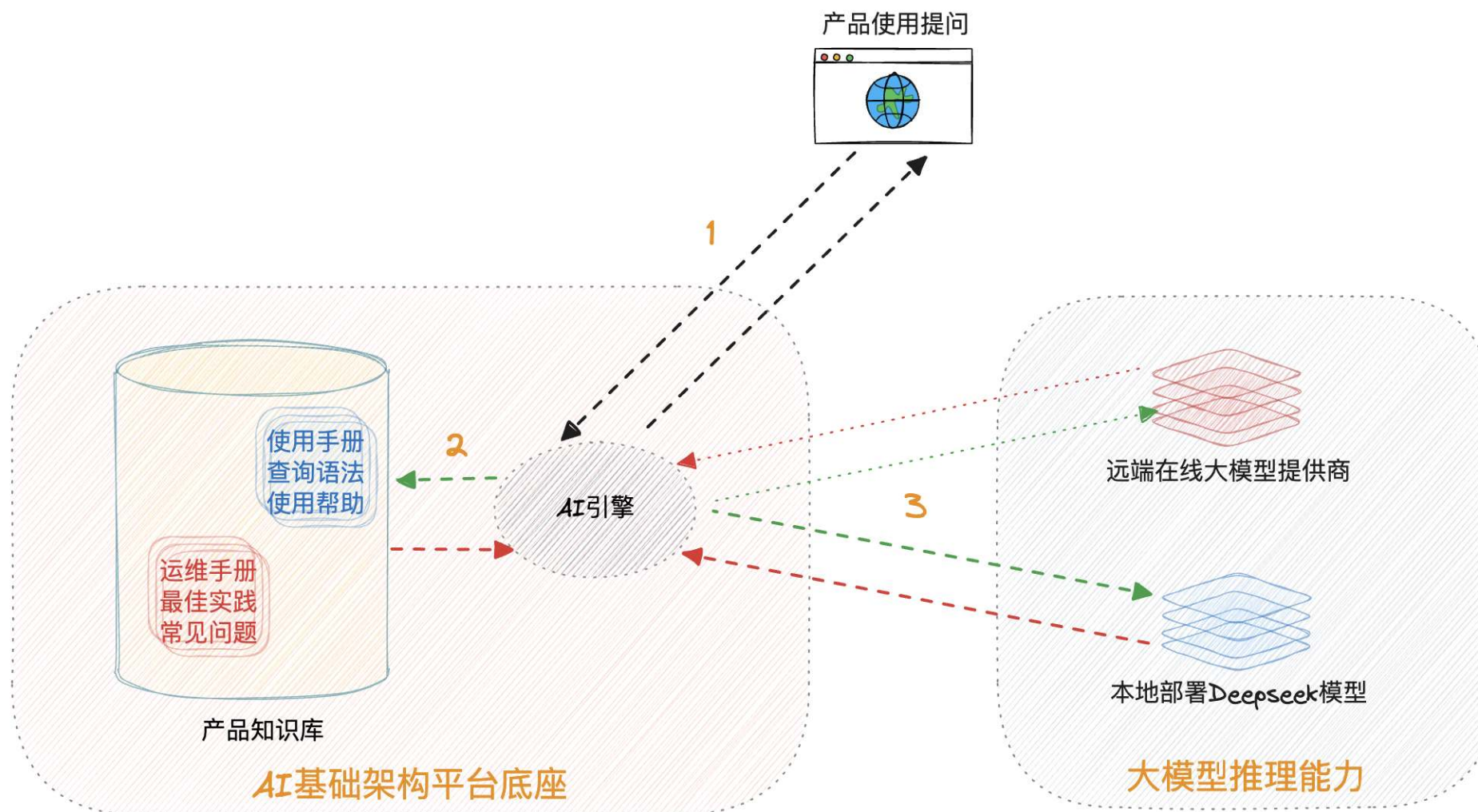
场景说明

- 需求：为炎凰产品提供动态问答能力，解决静态文档交互性差的问题
- 挑战：内网环境限制、文档版本频繁迭代、知识库规模较小（RAG优先于微调）

工程实践

- 自研工具使用：开发命令行工具 aim，实现：
 - Markdown按标题层级自动切分
 - 批量上传/删除/备份/恢复知识库文档
 - 版本备份与恢复

04 知识库问答架构图



04 知识库问答界面

 使用手册

 新对话

 New conversation

New conversation

什么是表函数?



表函数是一种在SQL中使用的特殊类型的函数，它返回一个数据行集合，即一个二维表结构。与普通的标量函数不同，标量函数返回的是一个单一的值。表函数可以看作是一个带参数的视图，它允许用户通过参数来过滤或转换数据，从而返回满足特定条件的行集。

在炎凰数据平台中，表函数类似于一个参数化的视图，用户可以通过它执行添加、修改、删除等操作。例如，在炎凰数据平台中，你可以通过以下步骤来管理和使用表函数：

1. 进入数据目录页面，通过左侧导航栏的“表函数”进入管理界面。
2. 在“表函数”页面，用户可以点击左上方的黄色按钮来新建一个表函数。
3. 对于系统内置的表函数，可能需要提供相应的配置项、插件等配置，例如对于JDBC表函数，用户可能需要上传JDBC driver和JDBC数据源配置。

总之，表函数在炎凰数据平台中是一个强大的工具，它允许用户以灵活和高效的方式处理和查询数据。

引用

 yhp_2-17-0_docs_3.pdf

04 典型场景二 自然语言转 SQL 场景

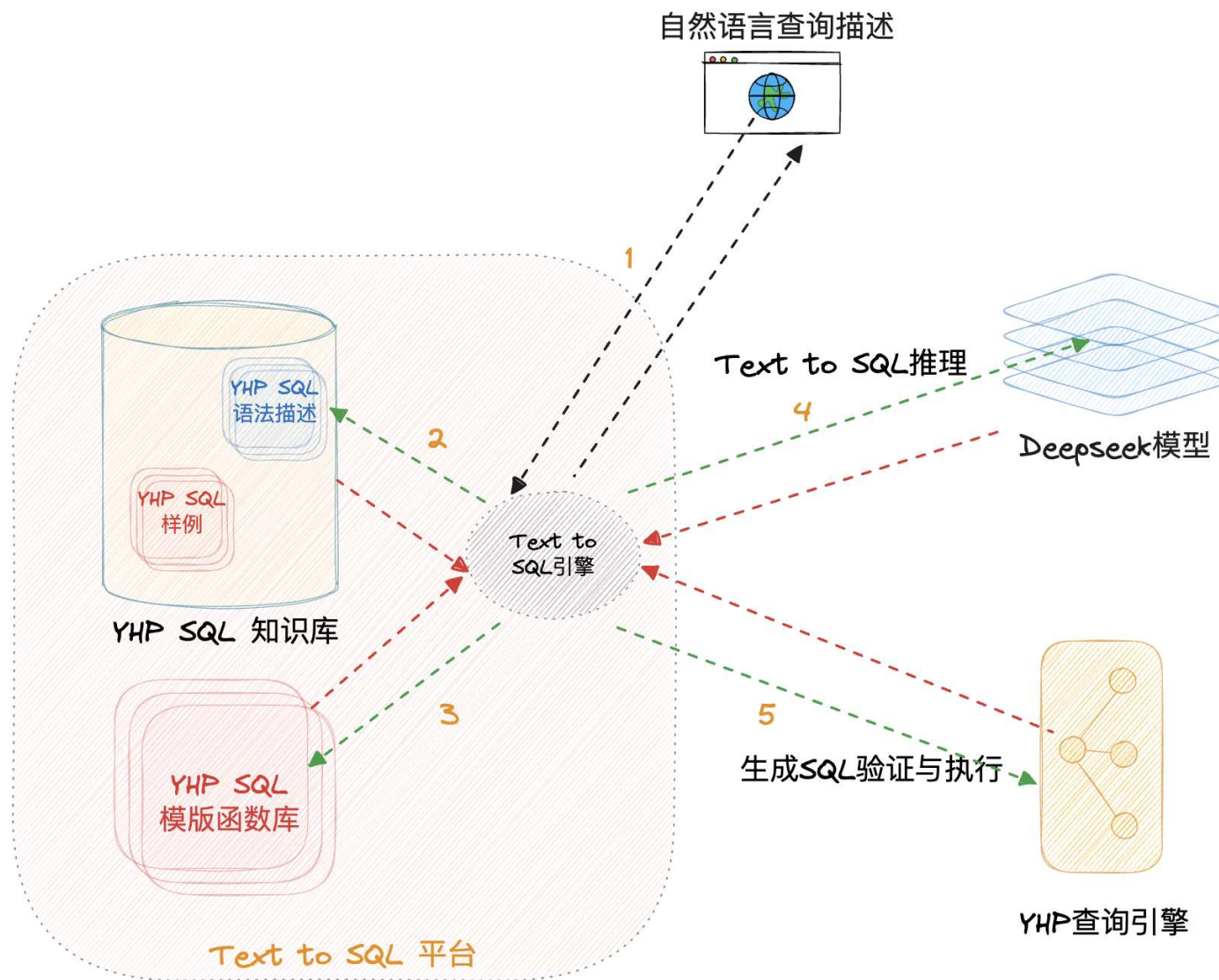
场景说明

- 目标：将自然语言查询转换为炎凰定制SQL（YHP SQL），降低用户使用门槛
- 技术选型：基于 Vanna框架 + DeepSeek模型，结合工具调用优化准确性

工程实践

- 自研工具使用：开发命令行工具 aim，实现：
 - YHP SQL 语法定义的训练数据准备
 - YHP SQL测试数据到样例数据的批量转换
- 工具调用优化：扩展Vanna开源版，添加YHP SQL模板函数

04 自然语言生成YHP SQL架构图



04 广泛的模型支持



DeepSeek Ready



模型供应商

OpenAI OpenAI 提供的模型，例如 GPT-3.5-Turbo 和 GPT-4。 LLM TEXT EMBEDDING SPEECH2TEXT MODERATION TTS	deepseek 深度求索提供的模型，例如 deepseek-chat、deepseek-coder。 LLM	ANTHROPIC Anthropic 的强大模型，例如 Claude 3。 LLM
Azure OpenAI Service LLM TEXT EMBEDDING SPEECH2TEXT TTS	Gemini 谷歌提供的 Gemini 模型。 LLM	Google Cloud Vertex AI in Google Cloud Platform. LLM TEXT EMBEDDING
NVIDIA API Catalog LLM TEXT EMBEDDING RERANK	NVIDIA NVIDIA NIM，一组易于使用的模型推理微服务。 LLM	cohere LLM TEXT EMBEDDING RERANK
upstage Upstage 提供的模型，例如 Solar-1-mini-chat。 LLM TEXT EMBEDDING	aws Bedrock AWS Bedrock's models. LLM TEXT EMBEDDING RERANK	together.ai LLM

模型供应商

NVIDIA TRITON INFERENCE SERVER LLM	智谱 AI LLM TEXT EMBEDDING	BAICHUAN AI LLM TEXT EMBEDDING
讯飞星火 LLM	MINIMAX LLM TEXT EMBEDDING	通义千问 LLM TTS TEXT EMBEDDING RERANK
文心一言 LLM TEXT EMBEDDING RERANK	Moonshot AI Moonshot 提供的模型，例如 moonshot-v1-8k、moonshot-v1-32k 和 moonshot-v1-128k。 LLM	Tencent Cloud SPEECH2TEXT
jina Embedding and Rerank Model Supported TEXT EMBEDDING RERANK	ChatGLM LLM	Yi 01.AI 零一万物提供的模型，例如 yi-34b-chat 和 yi-v1-plus。 LLM

05 Q&A

Q&A

感谢观看

<http://www.yanhuangdata.com>

