



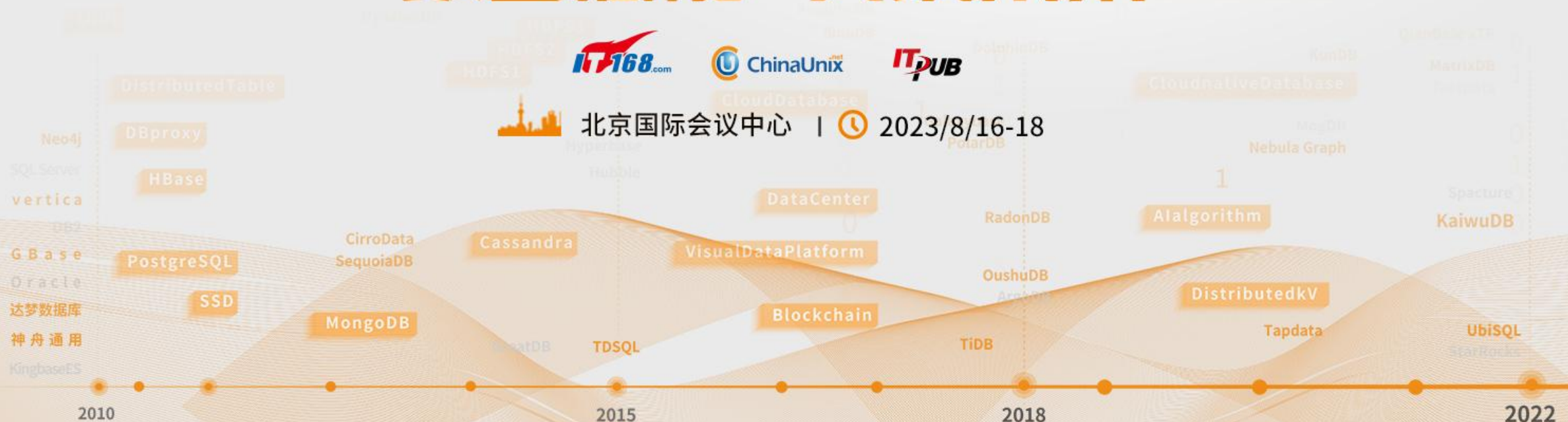
第十四届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA

数智赋能 共筑未来



北京国际会议中心 | 2023/8/16-18






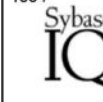


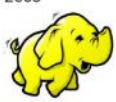











数据库智能化运维 实践分享

南京基石数据 CTO 白鳢

数字化转型的挑战

稳定增长的数据库种类

1977 	1983 	1984 	1989 	1992 	1994 
1995 	1996 	2005 	2009 	2009 	2010 
2012 	2014 	2015 	2016 	2017 	2018 

发展

大数据技术的应用

去“IOE” 风潮

虚拟化、云计算

国产化浪潮

容器技术的使用

促进

造成

成千上万种数据库



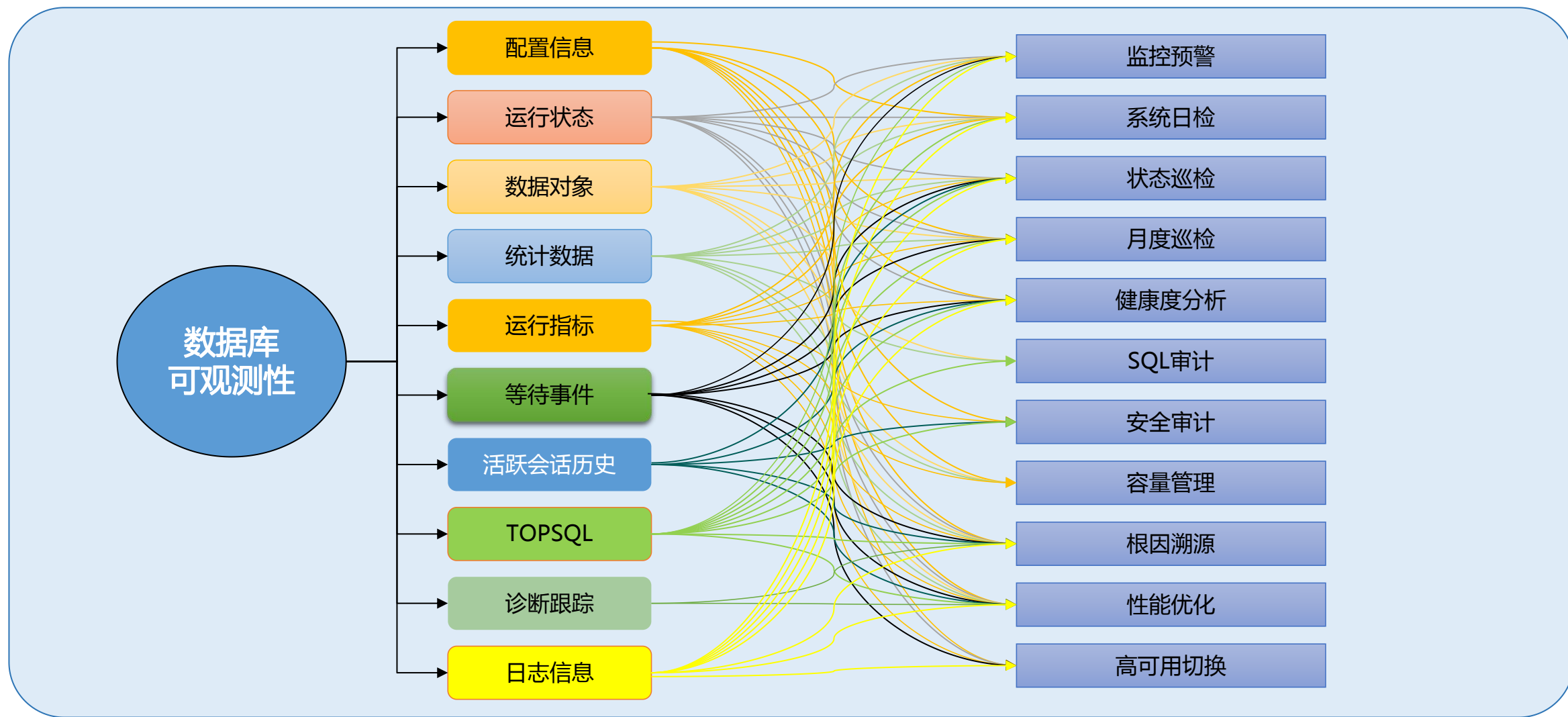
如何快速积累经验？ 如何高效运维？

DTCC 数智赋能 共筑未来

智能化运维的数字化基础

DTCC 2023

第十四届中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2023



关键建设工作

01

- 运维对象梳理
- 运维对象健康要点分析
- 运维对象指标集梳理
- 运维对象关键指标梳理
- 运维对象编码标准设计
- 指标分类标准设计
- 指标编码标准设计
- 指标标准属性设计
- 指标存储结构设计
- 指标分级存储规范设计
- 指标订阅与应用接口规范设计

指标采集原则

02

安全可靠

- 采集方式安全可靠，不会破坏生产系统
- 遵循国网数据与网络安全规范

最小影响

- 避免过多消耗生产系统的资源
- 避免带来生产系统故障隐患

避免重复

- 一次采集，多处使用，避免同一指标多次采集
- 只采集基础数据，采集后在I6000平台里加工，减少同类指标采集次数

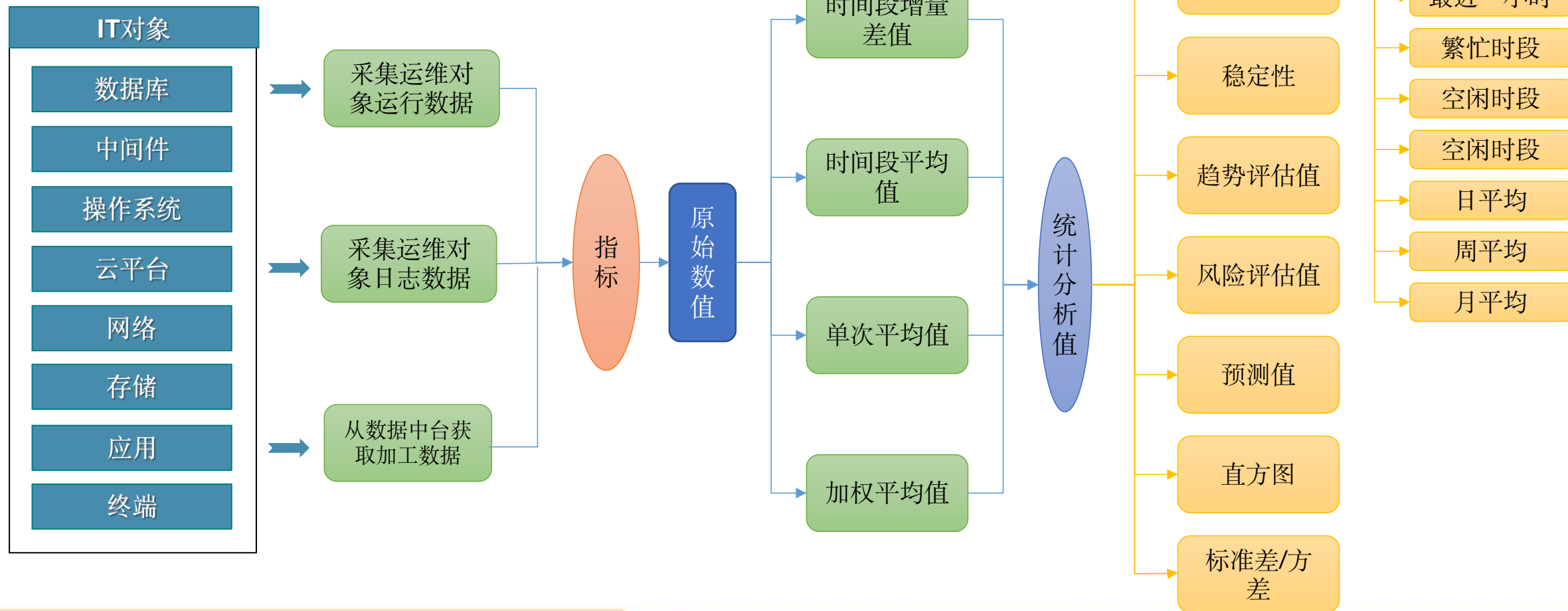
批量采集

- 一次连接采集多个指标，不允许单一指标建立独立连接采集
- 尽可能在一个命令中采集多个指标

指标加工

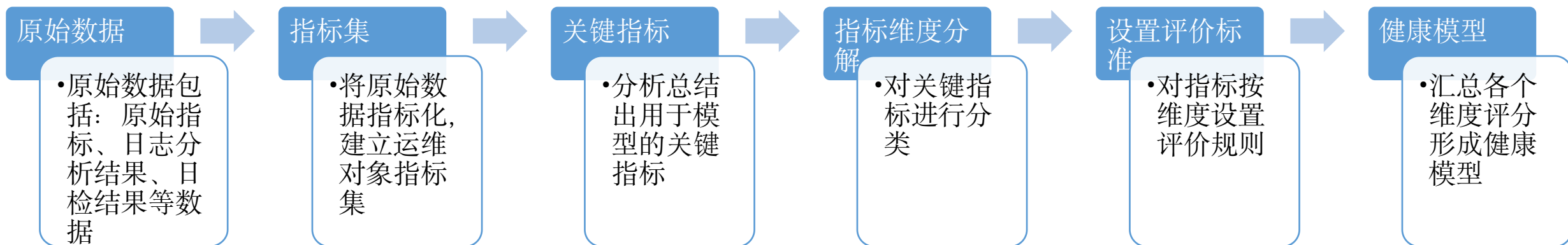
DTCC 2023

第十四届中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2023



运行状态模型描述运维对象的运行状态，包括健康模型、性能模型、负载模型、容量模型、故障模型，使用运维对象的指标数据来构建状态模型生成运行状态评价指标。模型分**专家模型**和**智能模型**，专家模型用于**评价打分**，智能模型则提供**智能评分及趋势预测**。

模型的关键指标筛选是通过专家经验结合深度学习算法的筛选完成的，其有效性已经在实际应用中得到了大量的验证。



第十四届中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2023

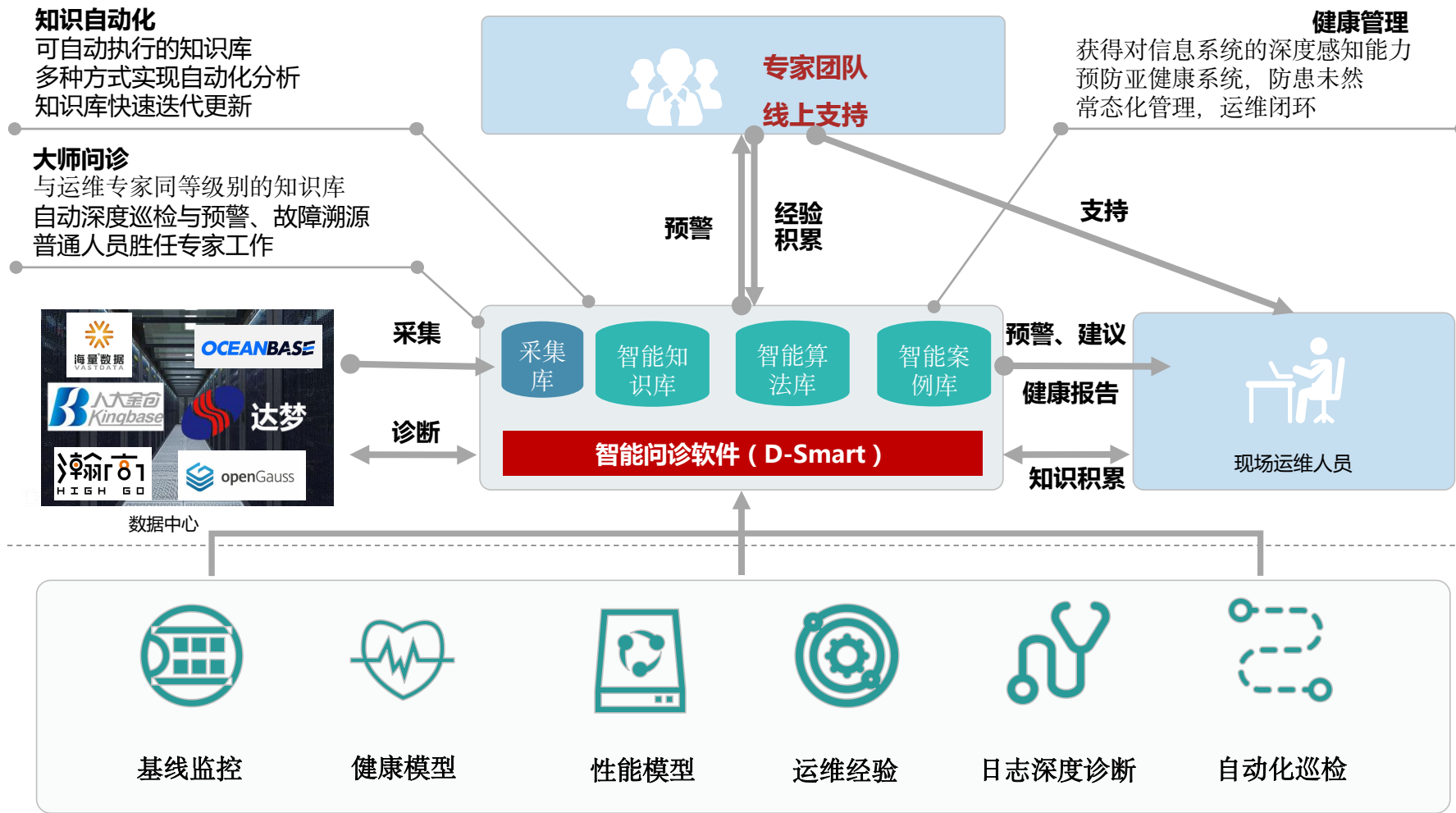
数字化模式下的运维体系重构

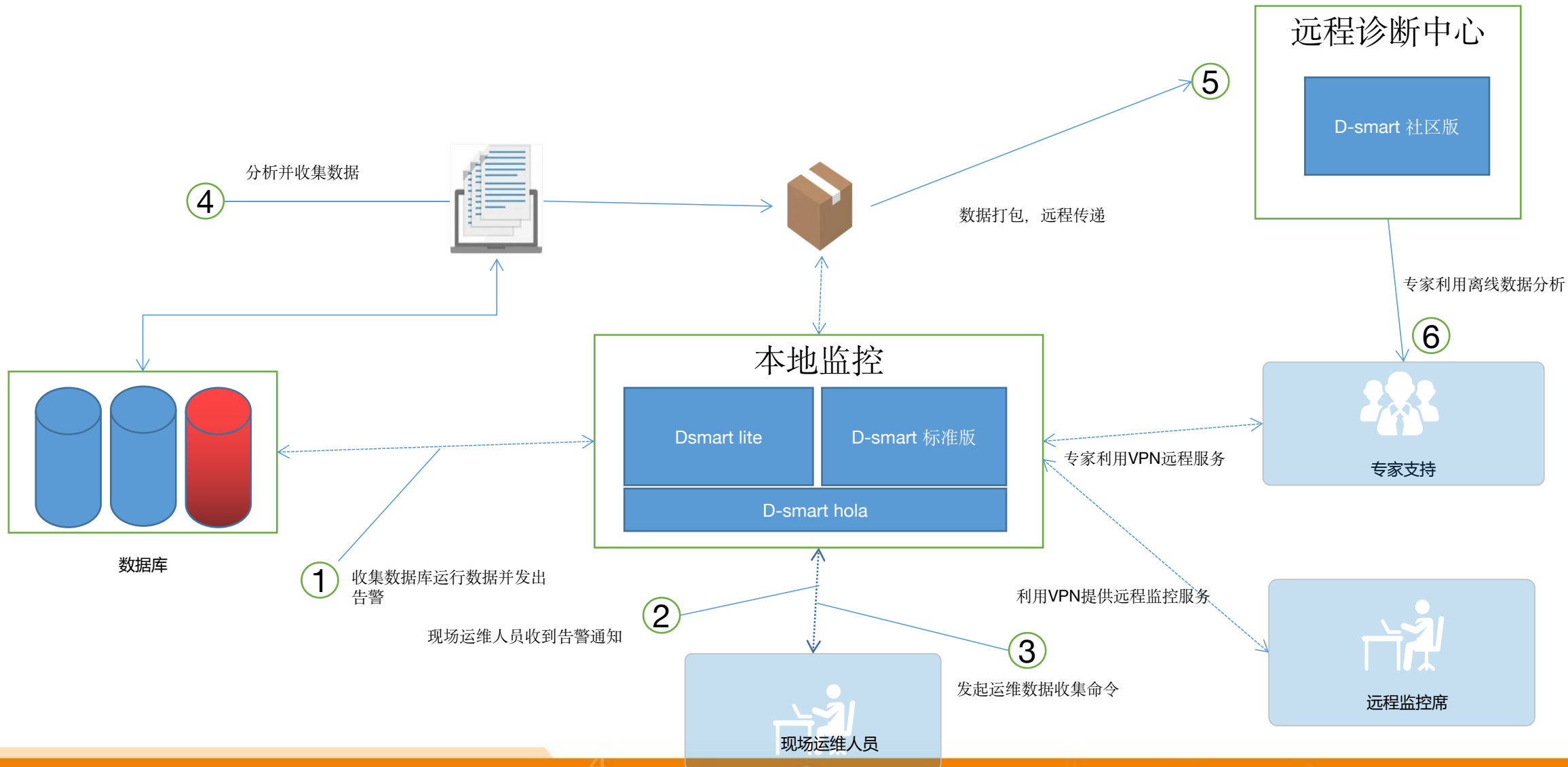
01

以“**知识自动化**”为核心理念实现“**运维数字化**”，以自动化分析与智能化分析替代人工。

02

运维人员利用平台，实时监控数据库运行状态并结合其丰富的运行指标采集进行**自动化、智能化分析**进行预警，智能定位问题，并告知用户如何快速解决故障，**实现运维闭环**。





5.1 等待事件分析： 折叠

运维诊断 - 运维对象名称:

IP地址: 60.60.60.123

IO_EVENT	DataFileWrite	2	0.019	0.01
IO_EVENT	DataFileSync	2	0.009	0.00

【诊断结论】当前系统的问题主要因素可能集中在以下几个方面:

1.应用类: 63.28%;

- > 1.1. 表连接操作.
- > 1.1.1. HASH JOIN操作.
- > 1.2. 存在Top SQL

2.并发类: 21.88%;

- > 2.1. 数据库逻辑读数量过大.
- > 2.2. 热块争用问题.
- > 2.3. IO并发量或者并发性能存在问题.
- > 2.4. 并发事务量过大.
- > 2.5. WAL/REDO并发, 数据库写并发过大.
- > 2.6. 硬盘排序过多.

3.配置类: 10.16%;

- > 3.1. 数据库配置存在问题.
- > 3.1.1. DBCACHE配置存在问题.
- > 3.1.2. 检查点配置存在问题.

【诊断结论】当前系统的问题次要因素可能集中在以下几个方面:

1.IO类: 3.91%;

- > 1.1. IO延时过大.
- > 1.1.1. 写IO延时大.

2.故障类: 0.78%;

- > 2.1. 数据库故障.
- > 2.1.1. 数据库出现过宕机或者重启现象.

平均等待时间(MS)



【问题】指定时间窗口存在平均等待时间较大的事件, 请参考上面提示的方向进行优化。

导出报告

保存报告

(2) 诊断项: HASH JOIN操作

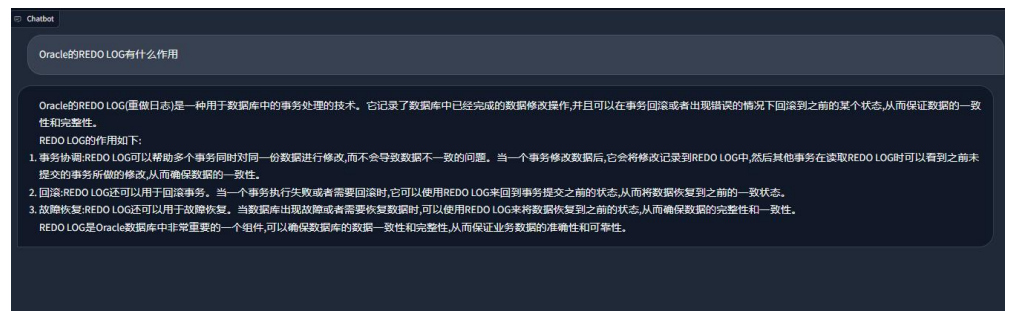
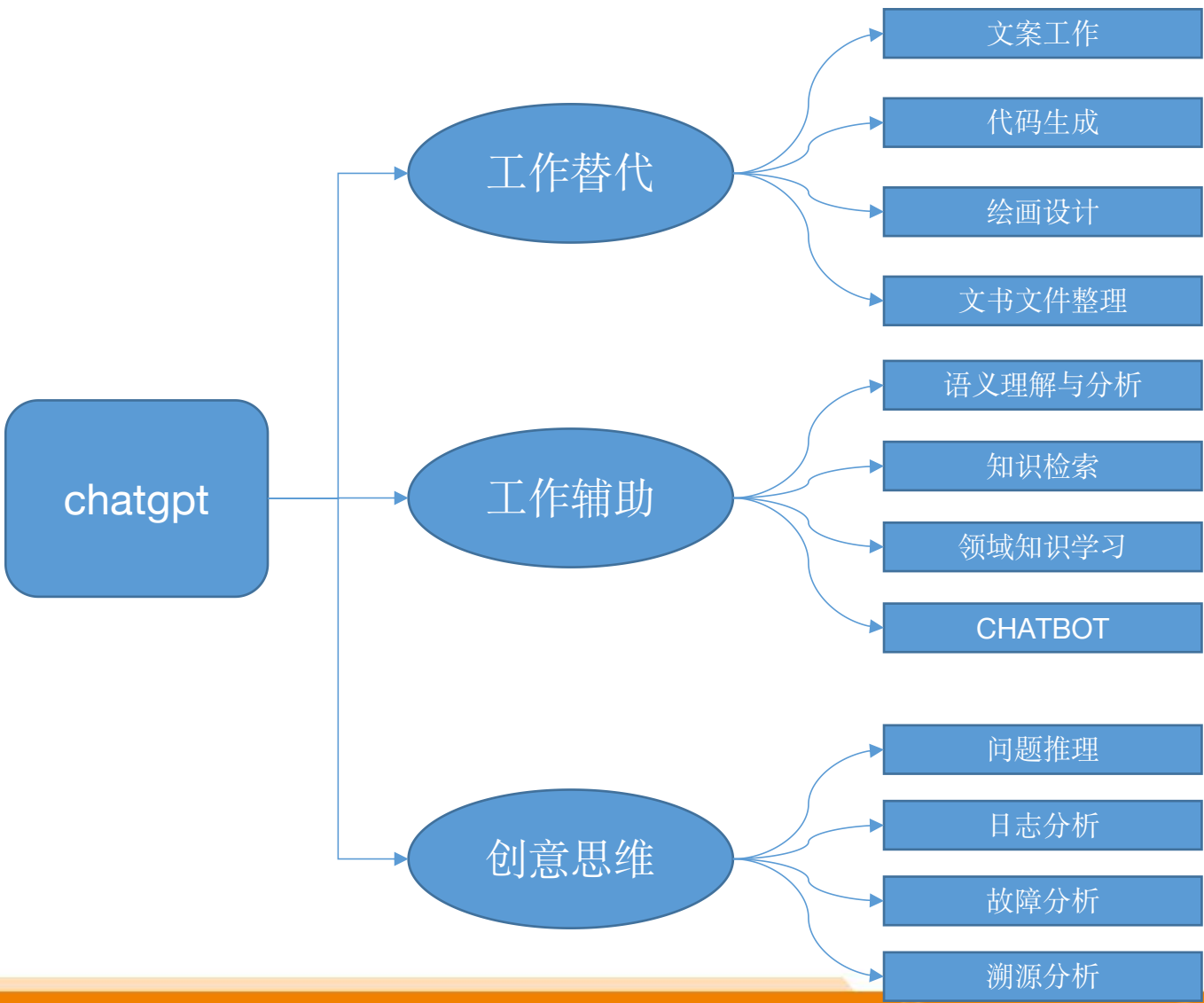
1. 逻辑读较多的TOP SQL
2. 物理读较多的TOP SQL
3. PostgreSQL Top SQL-按临时读写排序
4. 对PostgreSQL的Top SQL进行分析
5. 对PostgreSQL的Top SQL汇总进行分析
6. PostgreSQL Top SQL-按逻辑读排序
7. PostgreSQL Top SQL-按物理读排序
8. PostgreSQL Top SQL-按处理行数排序
9. PG核心指标分析
10. 泛路由-PG活跃会话数分析
11. 泛路由-CPU使用率高OS层面分析
12. 泛路由-IO问题操作系统层面分析
13. 泛路由-内存问题操作系统层面分析

(3) 诊断项: 存在Top SQL

1. 获取CPU时间比较高的TOP SQL
2. 获取执行时间比较高的TOP SQL
3. pg高IO消耗 top sql
4. PostgreSQL Top SQL-按执行时间排序
5. PostgreSQL长事务分析
6. PostgreSQL锁等待分析

(4) 诊断项: 数据库逻辑读数量过大

1. 数据库负载维度分析



大模型在数据库领域的应用场景

DTCC 2023

第十四届中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2023

知识库学习

- 知识问答
- 知识库管理
- 垂直搜索
- 智能案例库

诊断与推理

- 日志分析
- 故障诊断与分析

工作辅助

- SQL语句生成
- SQL优化辅助
- 数据库运行状态汇总分析
-

可用于数据库运维的大模型

DTCC 2023

第十四届中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2023

- 可用于数据库运维的大模型
 - 可本地化部署
 - 具有较强的编程能力
 - 中文能力较强
- 模型举例
 - CHATGLM-6B/CHATGLM—6B
 - 清华大学开源的大模型
 - 中文应答能力较强
 - 可进行PTUNING/FINE TUNE
 - VICUNA-13B
 - 基于LLAMA的优化模型
 - 编程能力较强

数据库运维领域大模型Prompt工程

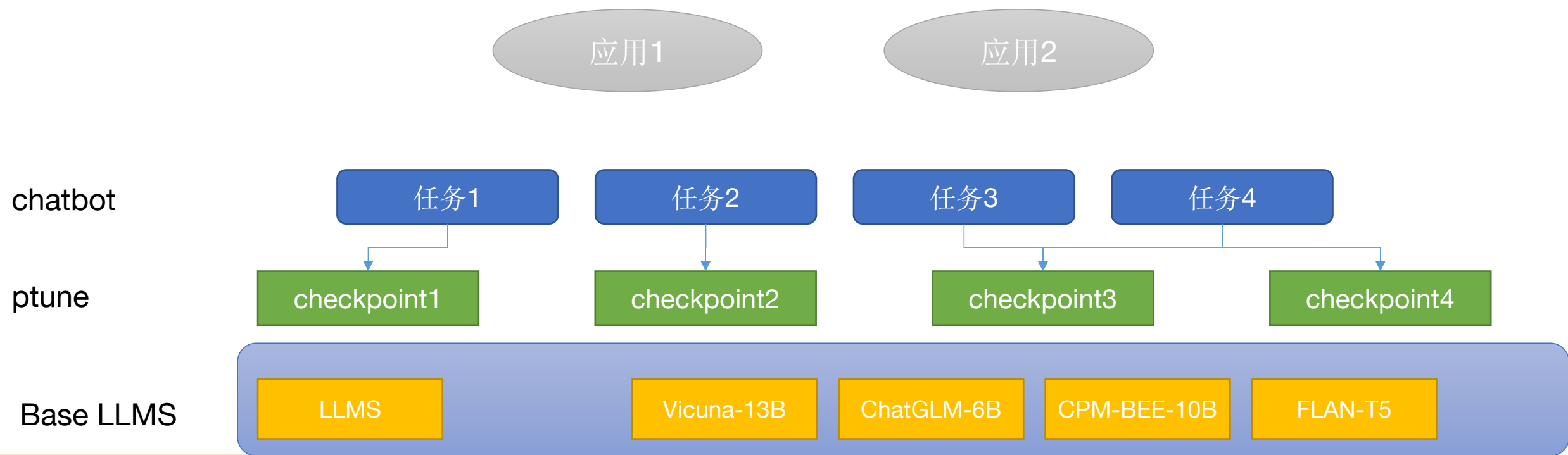
DTCC 2023

第十四届中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2023

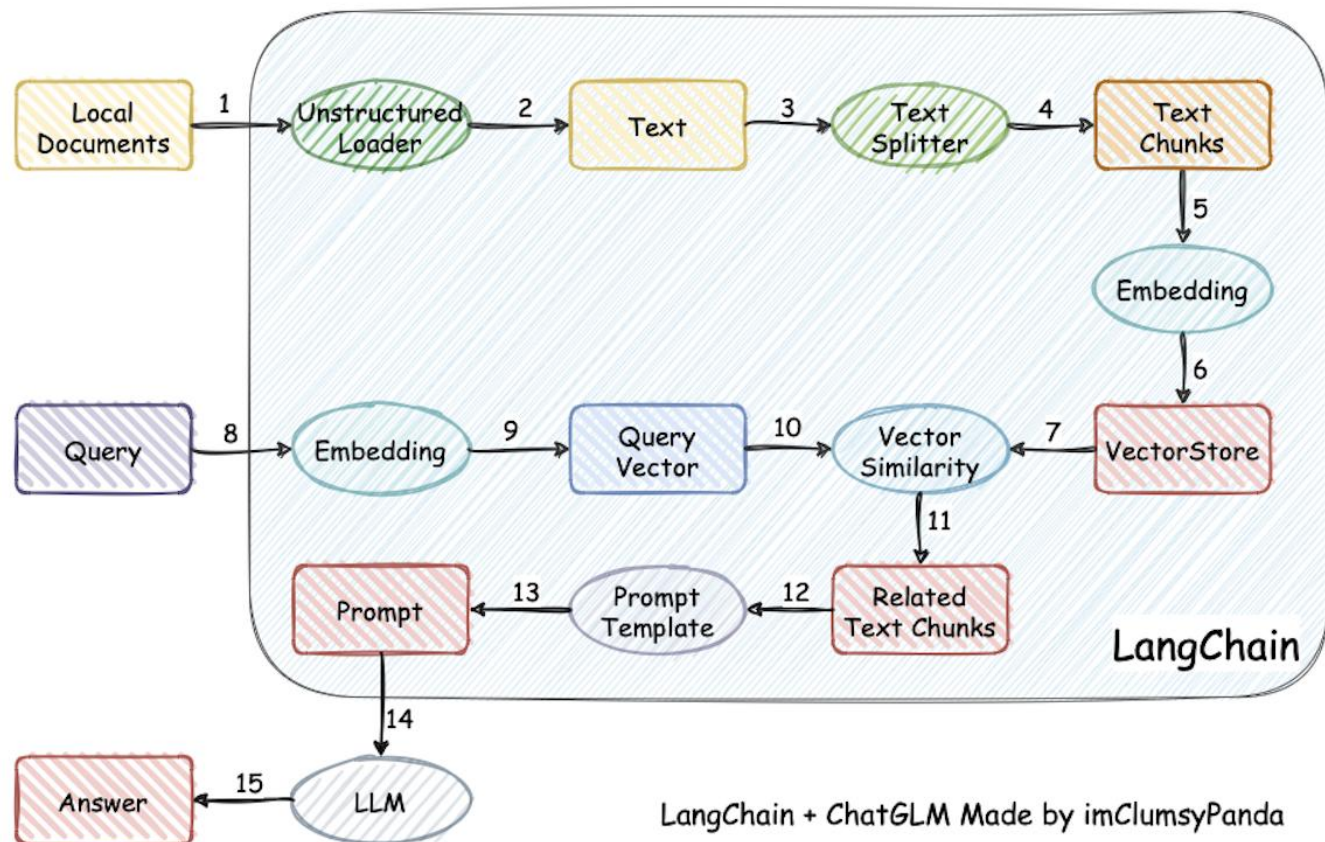
- Fine-tuning
 - 小数据集微调，效果显著
 - 微调数据集质量有要求
 - 训练成本较高
- Fixed-LM Prompt Tuning
 - 锁定大模型，通过提示数据微调模型
 - 适合few shot场景，训练成本较低
 - Zero shot任务性能较差
 - 微调训练集的质量很关键
- vectorDB本地知识库+autoprompt
 - 适合自然语言推理、事实检索、关系抽取等场景，实现无参数学习
 - 不需要存储模型的检查点，只需要保存生成的触发词和模板，就可以在现有的预训练语言模型上使用
 - 性能依赖于预训练模型的质量
 - 存在一定的局限性误差

Fixed-LM Prompt Tuning

- 对高质量的样本数量一定要求
- 容易产生过拟合，需要反复调参尝试
- 存在一定的随机不确定性
- 每个PT最好针对某个特定的任务



LLM+LANGCHAIN+矢量化本地知识库



• 工作流程

- 对本地知识库进行文本分割处理
- 通过适量查询精确定位知识
- 通过AutoPrompt生成对话模板
- 对LLM进行提问产生回答

• 注意事项

- 基础大模型的质量决定了最终应答质量
- 知识分类单独建库可以提高回答质量

- 知识图谱
 - 优点
 - 结构化知识
 - 准确性
 - 可解释性
 - 可持续发展
 - 缺点
 - 不完整
 - 缺乏语言理解能力
 - 看不见事实
- LLM
 - 优点
 - 语言处理与推理
 - 通用化能力强
 - 适用面广泛
 - 缺点
 - 隐性知识
 - 幻觉问题
 - 不可知的黑盒
 - 较难用于决策
 - 领域知识/新知识更新问题
- 如果用知识图谱增强 LLM，那么知识图谱不仅能被集成到 LLM 的预训练和推理阶段，从而用来提供外部知识，还能被用来分析 LLM 以提供可解释性。
- 用 LLM 来增强知识图谱，比如知识图谱嵌入、知识图谱补全、知识图谱构建、知识图谱到文本的生成、知识图谱问答。LLM 能够提升知识图谱的性能。
- 知识图谱与LLM协同：利用知识图谱进行初期推理，用结果生成PROMPT，利用LLM问答获得最终结果。

THANKS

TDDL

DistributedTable

DBproxy

HBase

PostgreSQL

SSD

MongoDB

GreatDB

Cassandra

Hyperbase

Hubble

DataCenter

VisualDataPlatform

Blockchain

ArgoDB

Distributed

DatabaseKernel

TemporalData

CloudnativeData

AIalgorithm