DTCC

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

IT168.com ChinaUnix ITPUB

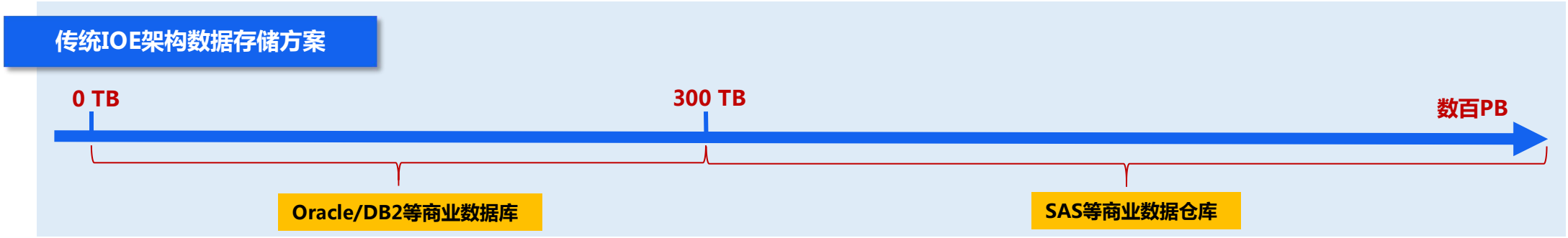
北京国际会议中心 | 2023/8/16-18

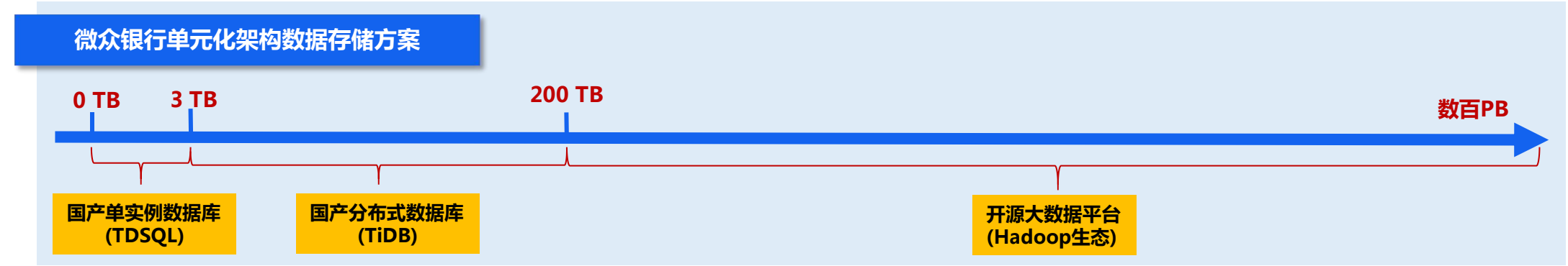
微众银行国产数据库应用实践

胡盼盼 / 微众银行数据库平台负责人

1. 数据库整体架构

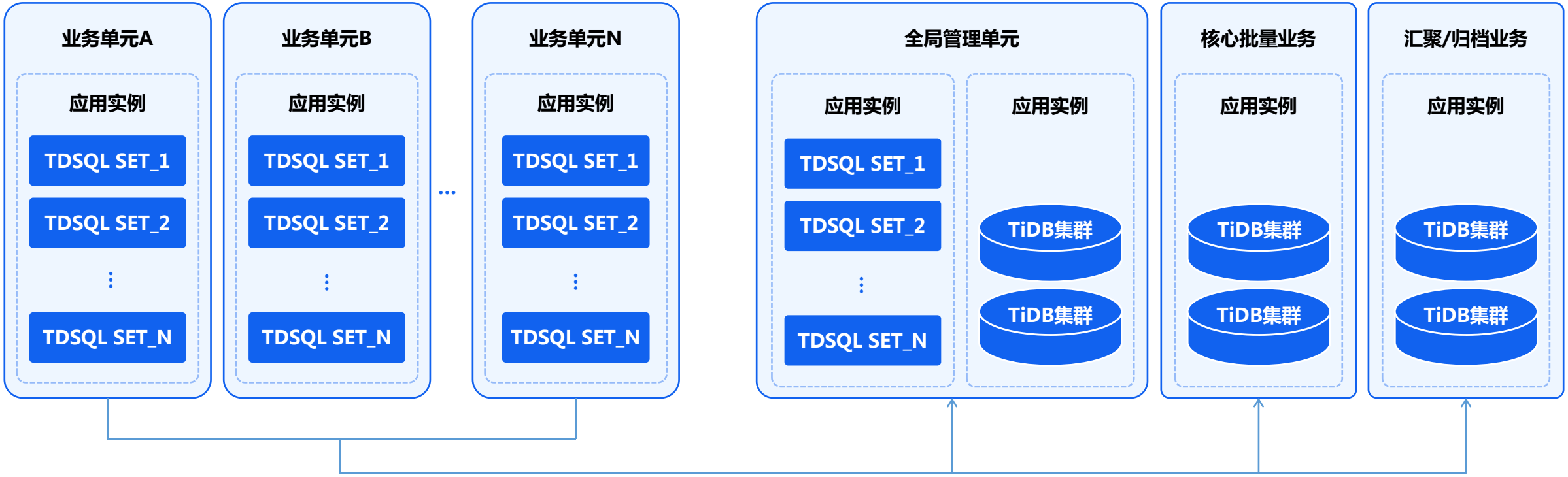
传统IOE架构与微众数据存储架构方案对比





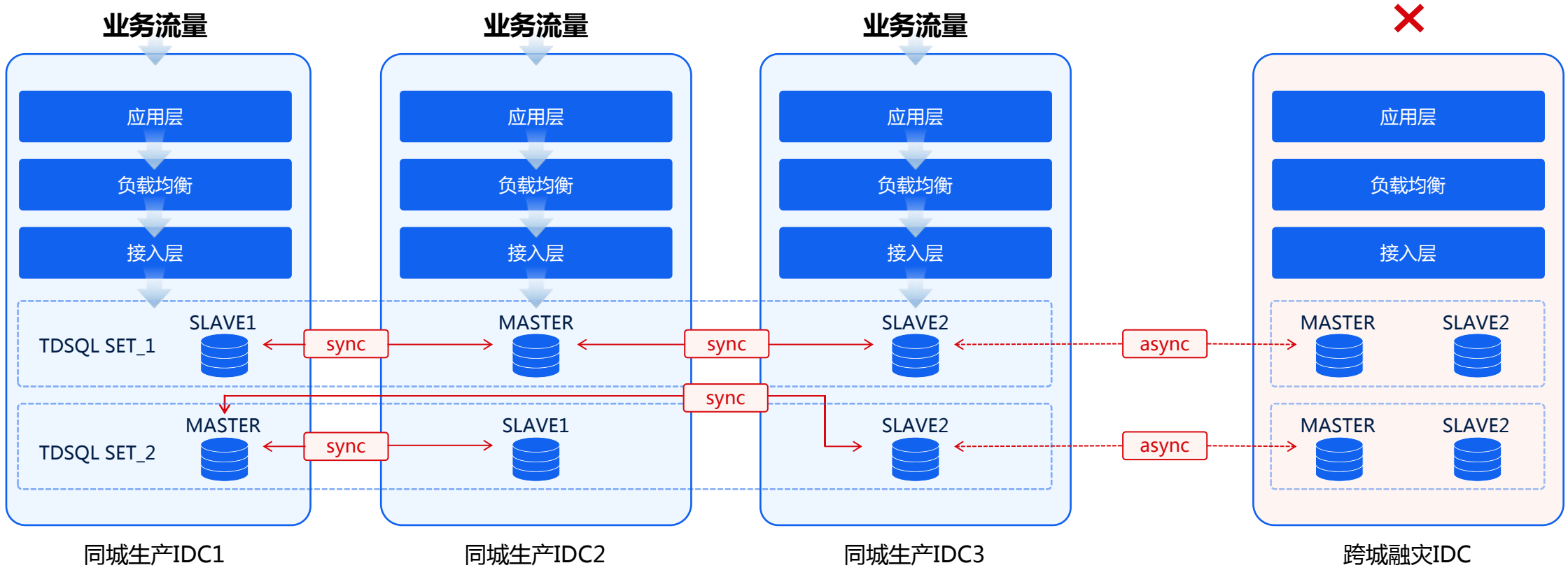
数据库整体架构方案

- 基于用户划分的单元化的分布式核心架构，每个业务单元形成自包含单位
- 业务单元和小容量全局管理单元业务（3TB以内）采用TDSQL单实例架构数据库；
- 大容量全局管理单元业务（3TB以上）、核心批量业务、汇聚/归档类业务采用TiDB分布式数据库



TDSQL同城多活部署(TiDB同架构部署)

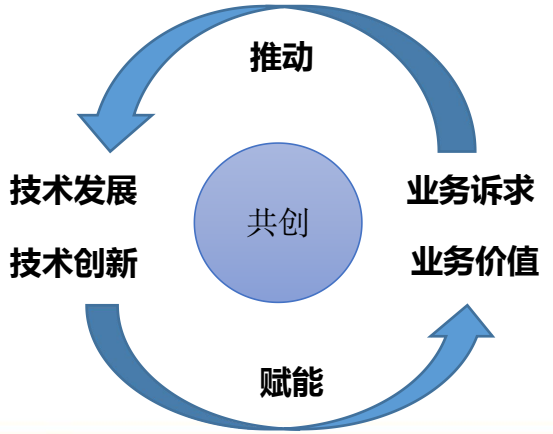
3(同城副本) + 2(跨城副本) 的TDSQL部署架构，同城主备强一致性数据同步， RPO=0；RTO <= 30秒



TiDB应用场景演进

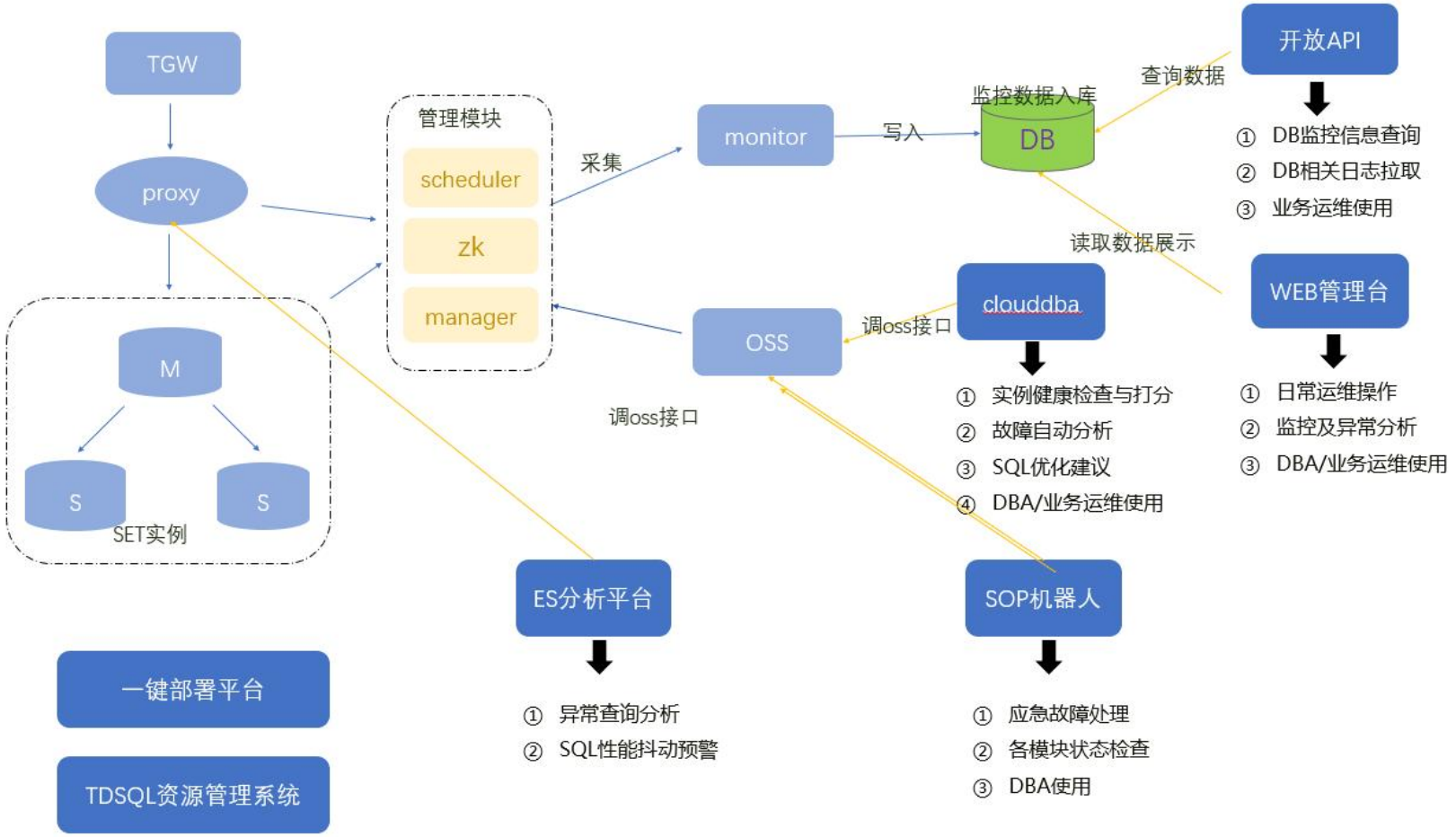
版本演进	TiDB 2.0	TiDB 3.0	TiDB 4.0	TiDB 5.4	TiDB 6.5	TiDB 7.0
TiDB特性	水平扩展性	性能提升 支持高并发、高吞吐	DM 2.0 乐观DDL 按需过滤event 热点预打散	大容量集群支持(200T) Compaction小步快走 HTAP TiFlash（列式引擎） 分区表GA	DM 6.5 解析效率提升 兼容MariaDB时间精度 悲观锁+RC 更可靠 TiCDC更可靠	Resource control Partitioned raft kv
业务价值	解决ADM单机瓶颈	联机中台、管理平台， 提供统一服务	多DCN汇聚 企业贷款运营管理台 贷款核心批量提效	HBase替换 HTAP 实时洞察、实时业务监控、广告实时归因	账务准实时流式批量处理	

50+业务系统，600+服务器节点

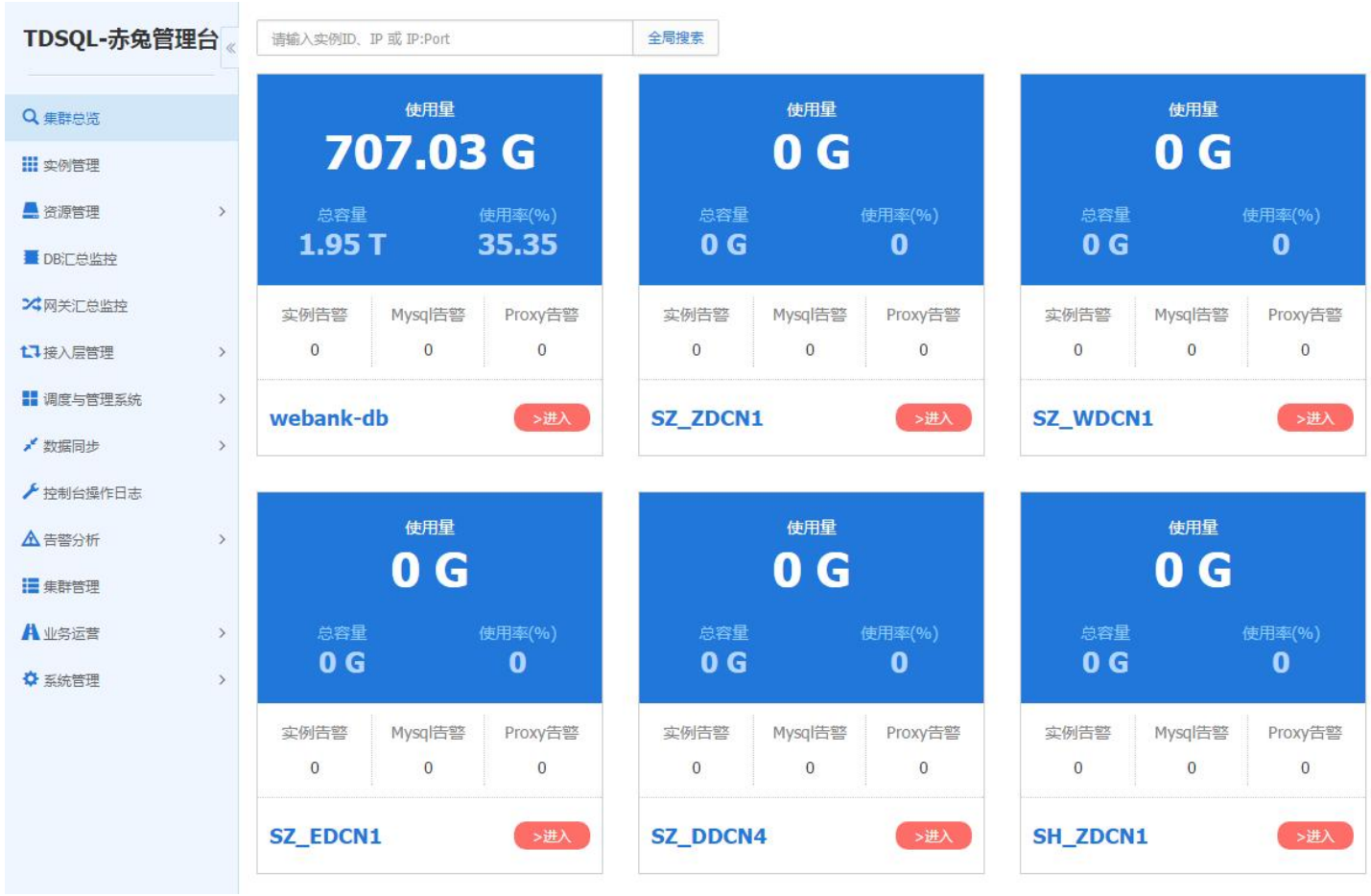


3.TDSQL运维实践

TDSQL运维体系架构

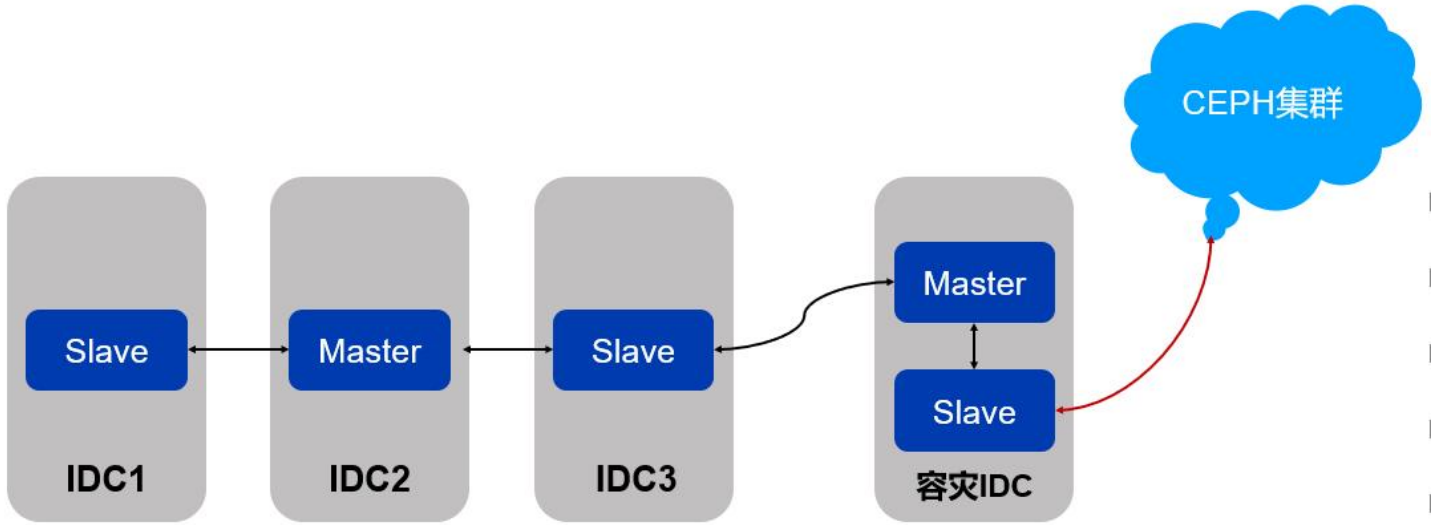


TDSQL管控平台



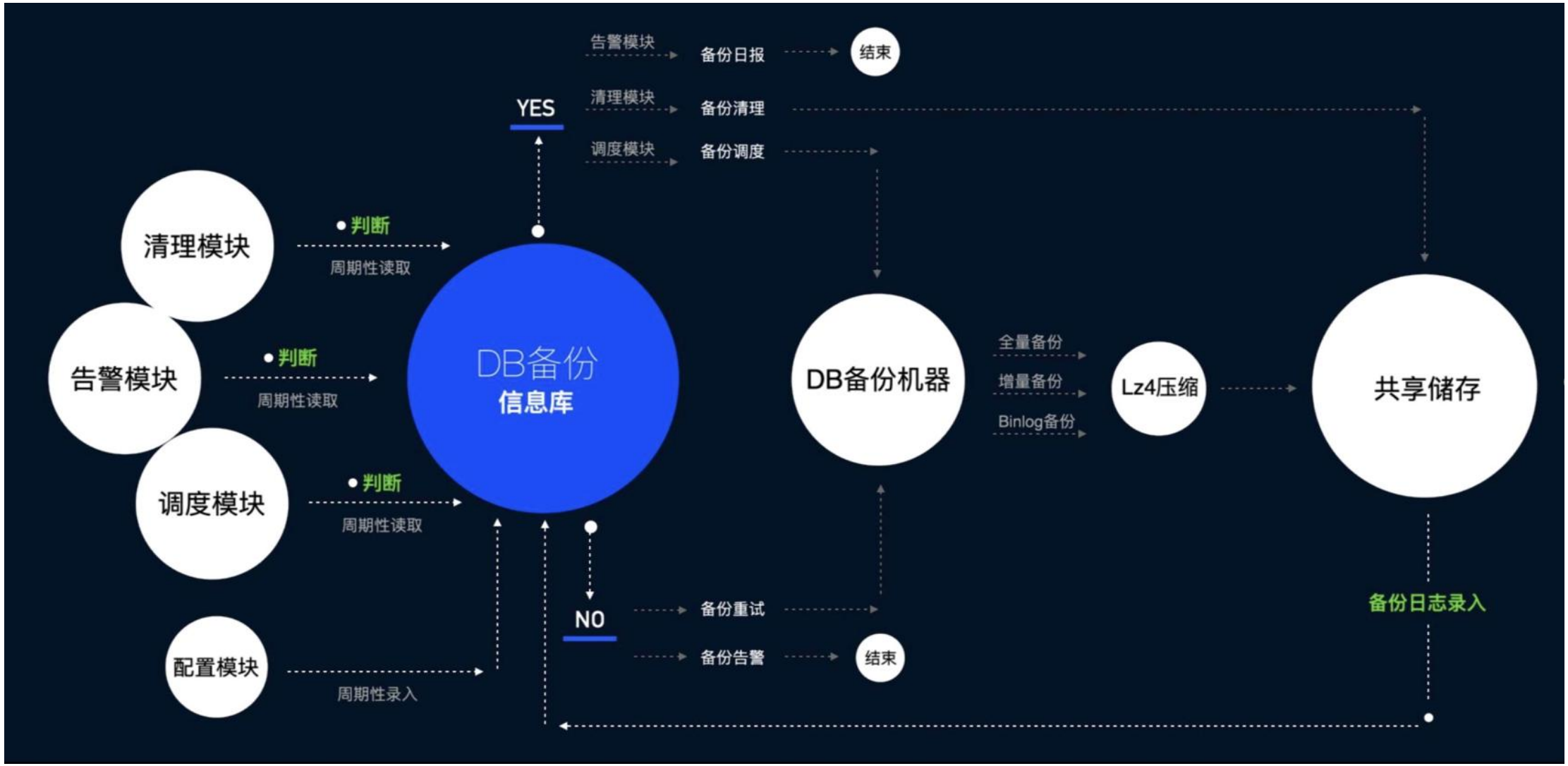
- 监控数据采集与上报
- 告警与策略配置
- 日常运维操作
- 慢查询分析与报表
- 性能分析与报表

基于APScheduler自建TDSQL备份平台

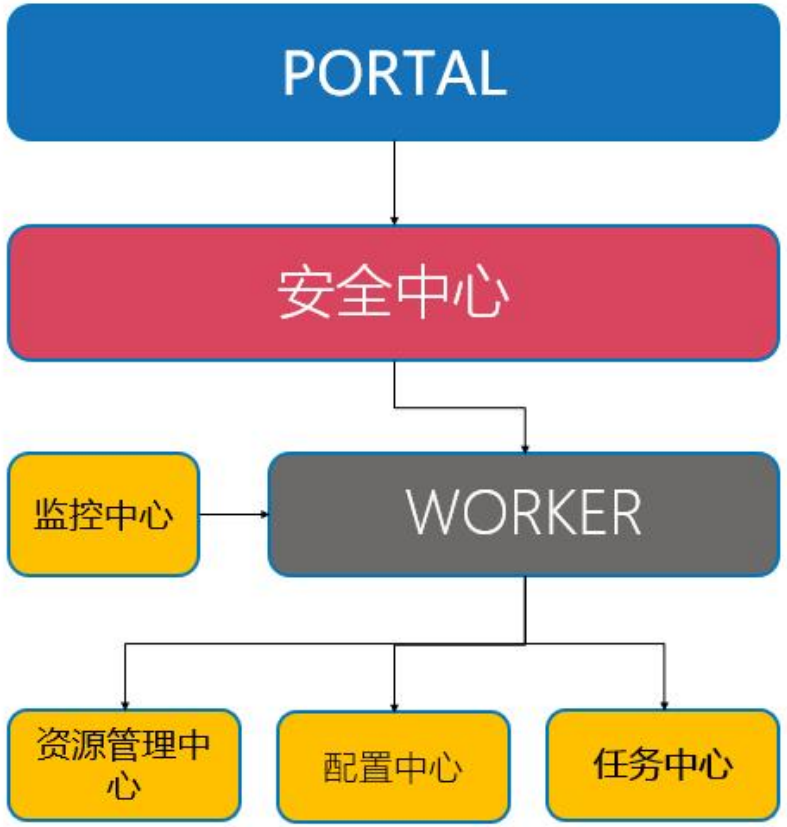


- 自研备份任务调度系统
- CEPH集群保存所有备份文件
- 全量物理备份,少量逻辑备份
- 每周日全备, 每天增备
- Binlog 5分钟实时备份
- 三个月一次备份恢复演练

TDSQL备份系统流程图



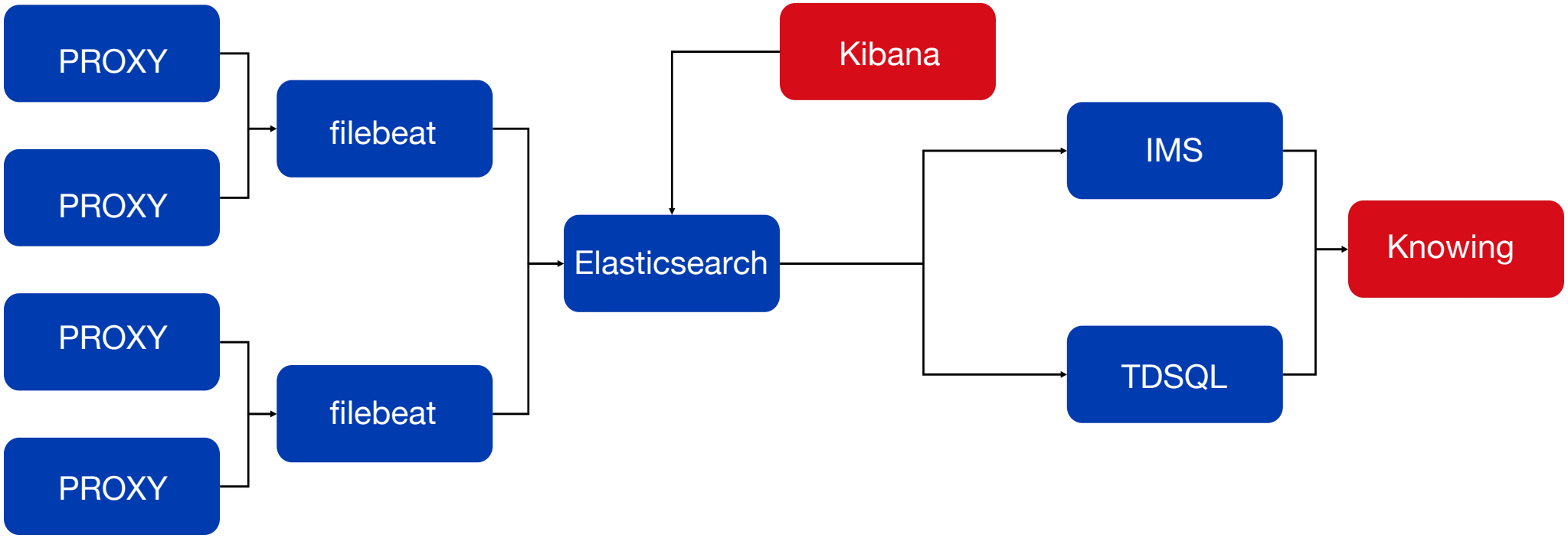
TDSQL准生产数据恢复平台



TDSQL自动恢复数据平台架构

- PORTAL: 提供开放接口。包括提交恢复任务，查询任务进度，取消任务等。
- 安全中心: 验证用户权限、确保数据安全。
- WORKER: 执行恢复任务。当有多个恢复任务时，会唤起多个worker执行协作进行恢复操作。
- 资源管理中心:
  - ✓ 实时监控CEPH存储及网卡流量，动态调节恢复效率
  - ✓ 自动优先级队列。同批次的任务多个维度确认优先级
  - ✓ 管理准生产资源，提高资源利用率及恢复成功率
  - ✓ 搜寻备份信息，获取合适的备份集
- 监控中心: 监控worker的工作状态及恢复任务进度。自动确认失败原因，尽可能重新拉起失败任务
- 配置中心: 记录关键及个性化配置，一次配置，永久使用。
- 任务中心: 负责分发恢复任务、记录恢复任务及其相关状态信息

基于Elasticsearch全量SQL采集与分析平台



- 使用filebeat采集TDSQL proxy日志，切分清洗后存入ES
- IMS侧从ES中生成各种监控指标，如SQL耗时曲线等
- knowing侧从IMS获取Metrix指标，进行机器学习的训练与预测。根据SQL曲线趋势进行异常根因分析以及异常预警

TDSQL SQL预警分析规则

数据场景	数据粒度	匹配规则	举例	识别时效	告警频率	适应曲线
高频SQL	10分钟	当前采集点平均耗时环比最近7个采集点平均耗时上升50%	20ms->30ms	30分钟	即时触发	持续、稳定、波动小
高频SQL	10分钟	当日平均耗时同比最近7日平均耗时上升50% 当日平均耗时同比上周同期平均耗时上升50%	20ms->30ms	30分钟	即时触发	
高频SQL	天粒度	7日平均耗时同比30日平均耗时上升50%	20ms->30ms	3天	即时触发	
高频SQL	天粒度	7日平均耗时同比60日平均耗时上升100%	20ms->40ms	3天	即时触发	
关键SQL	10分钟	最近1个采集点平均耗时>阈值(通过界面设定)	5ms->100ms	10分钟	即时触发	
低频SQL	天粒度	7日平均耗时同比30日平均耗时上升50%	20ms->30ms	7天	即时触发	
低频SQL	天粒度	7日平均耗时同比60日平均耗时上升100%	20ms->40ms	7天	即时触发	

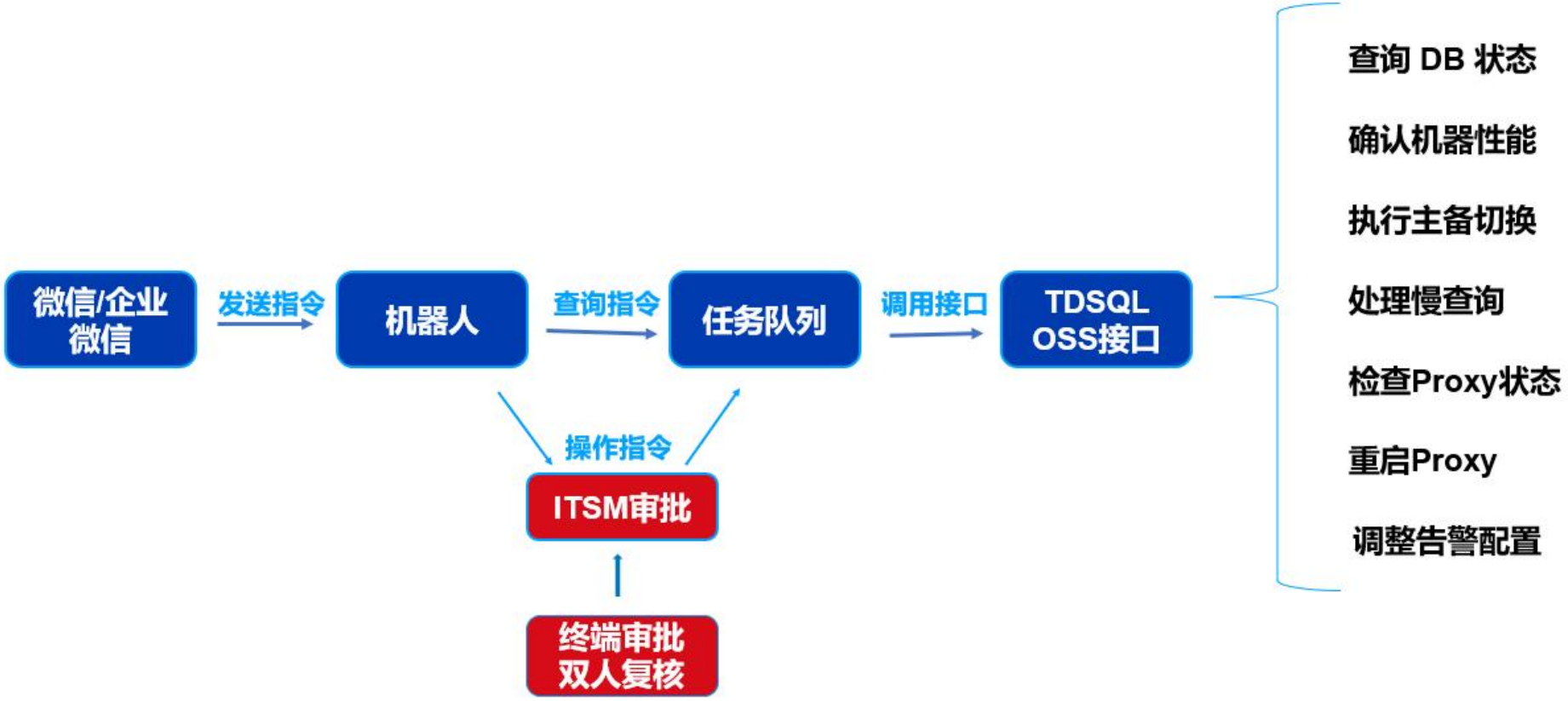
数据同步规则	高频、关键sql: 10分钟同步一次. 同步后本地存储一份。 全量sql: 一天同步一次。
高频SQL打标规则:	1. 初始标记： 日执行次数>1000 2. 更新活跃值： 10分钟同步一次高频，最新执行次数-上次执行次数>0， 日活跃数+1 3. 连续7天，日活跃数< 24*60*0.5 ， 取高频标识。
曲线特征打标规则	稳定曲线：最近30天，其中20天日平均耗时对比峰值耗时偏差<10%
报表	关键sql耗时报表
sql查询	查询： set_id, dbname, 子系统名称,sql关键字 展示: 高频sql展示10分钟粒度耗时曲线。 低频sql展示天维度平均耗时曲线。 编辑： 可以标识sql是否是关键sql

TDSQL应急预案

	TDSQL系统应急...	TDSQL(TDSQ... C	已审核	数据库运维室故障S...	murtonhuang(黄...	murtonhuang(黄...	计划时间 12-31 实际时间 未更新	计划时间 12- 实际时间 演练

场景名称	变更单	关联演练实例	场景更新计划	演练计划
TDSQL-故障-SET自动切换	[10070130]TDSQL-故障-SET主备...	不演练无需关联演练实例	计划时间 12-31 实际时间 未更新	计划时间 不演练
TDSQL-故障-SET 单备节点故障	[10072497]TDSQL-故障-SET备节...	不演练无需关联演练实例	计划时间 12-31 实际时间 未更新	计划时间 不演练
TDSQL-故障-自动切换失败处理...	[10077988]TDSQL-故障-自动切换...	[1]TDSQL-故障-自动切换失败处理流程	计划时间 12-31 实际时间 未更新	计划时间 03-01 实际时间 04-10 11:20 已演练
TDSQL-故障-网关节点故障	[10079018]TDSQL-故障-网关节点...	[4]TDSQL-故障-网关节点故障_演练	计划时间 12-31 实际时间 未更新	计划时间 04-01 实际时间 04-23 14:33 已演练
TDSQL-故障-zookeeper单节点...	[10074913]TDSQL-故障-zookeep...	不演练无需关联演练实例	计划时间 12-31 实际时间 未更新	计划时间 不演练
TDSQL-故障-scheduler节点故障	[10066274]TDSQL-故障-schedul...	不演练无需关联演练实例	计划时间 12-31 实际时间 未更新	计划时间 不演练
TDSQL-故障-两个备节点故障	请 提交 或 关联 变更单后演练	请演练后关联演练实例	计划时间 12-31 实际时间 未更新	计划时间 05-06 实际时间 已超时
TDSQL-binlog日志数据恢复	请 提交 或 关联 变更单后演练	请演练后关联演练实例	计划时间 12-31 实际时间 未更新	计划时间 06-01 实际时间 未演练
TDSQL-故障-冷备数据恢复	请 提交 或 关联 变更单后演练	请演练后关联演练实例	计划时间 12-31 实际时间 未更新	计划时间 07-01 实际时间 未演练
TDSQL-故障-zookeeper数据恢...	[10061155]TDSQL-故障-zookeep...	请演练后关联演练实例	计划时间 12-31 实际时间 未更新	计划时间 08-01 实际时间 未演练

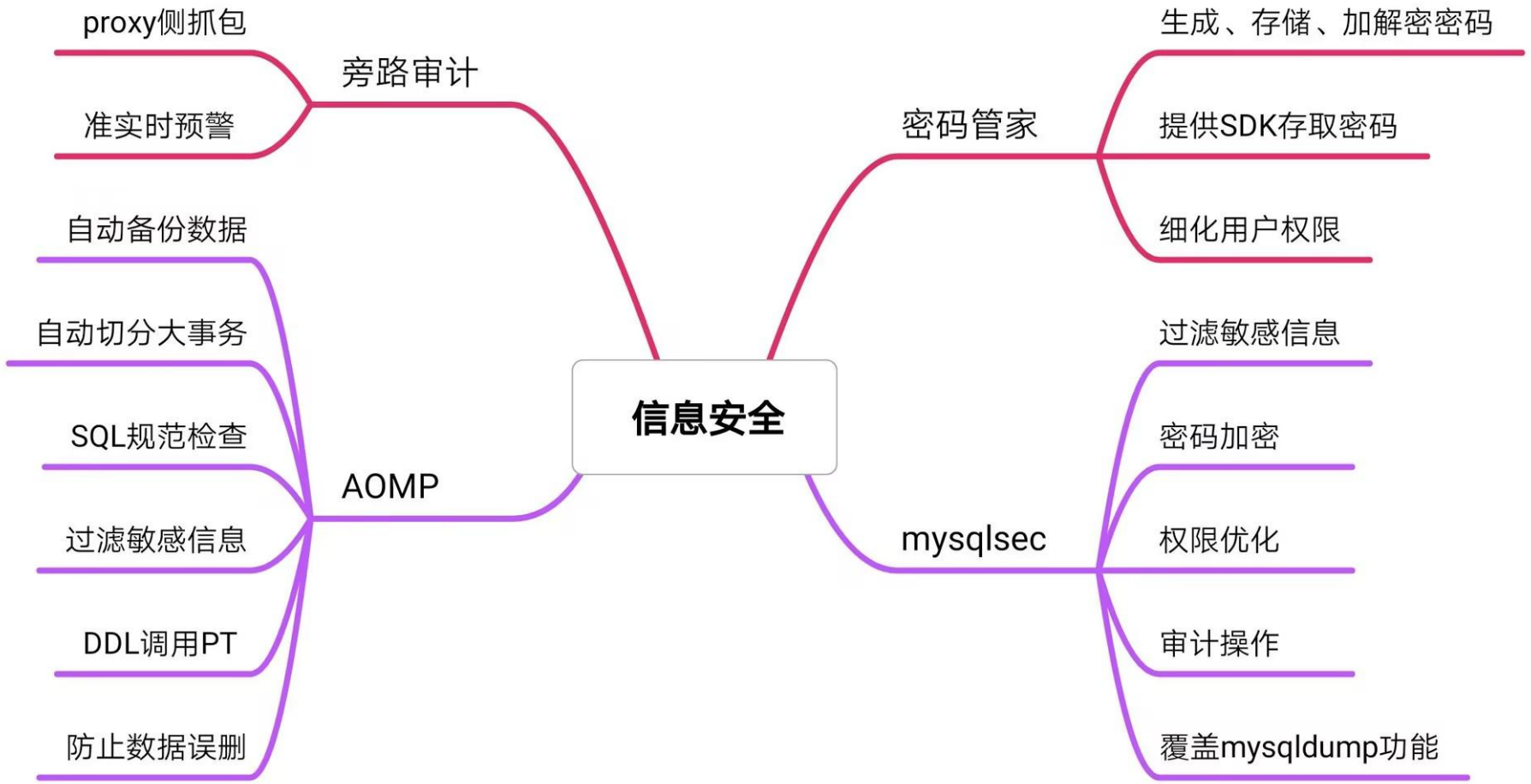
TDSQL自动化SOP机器人



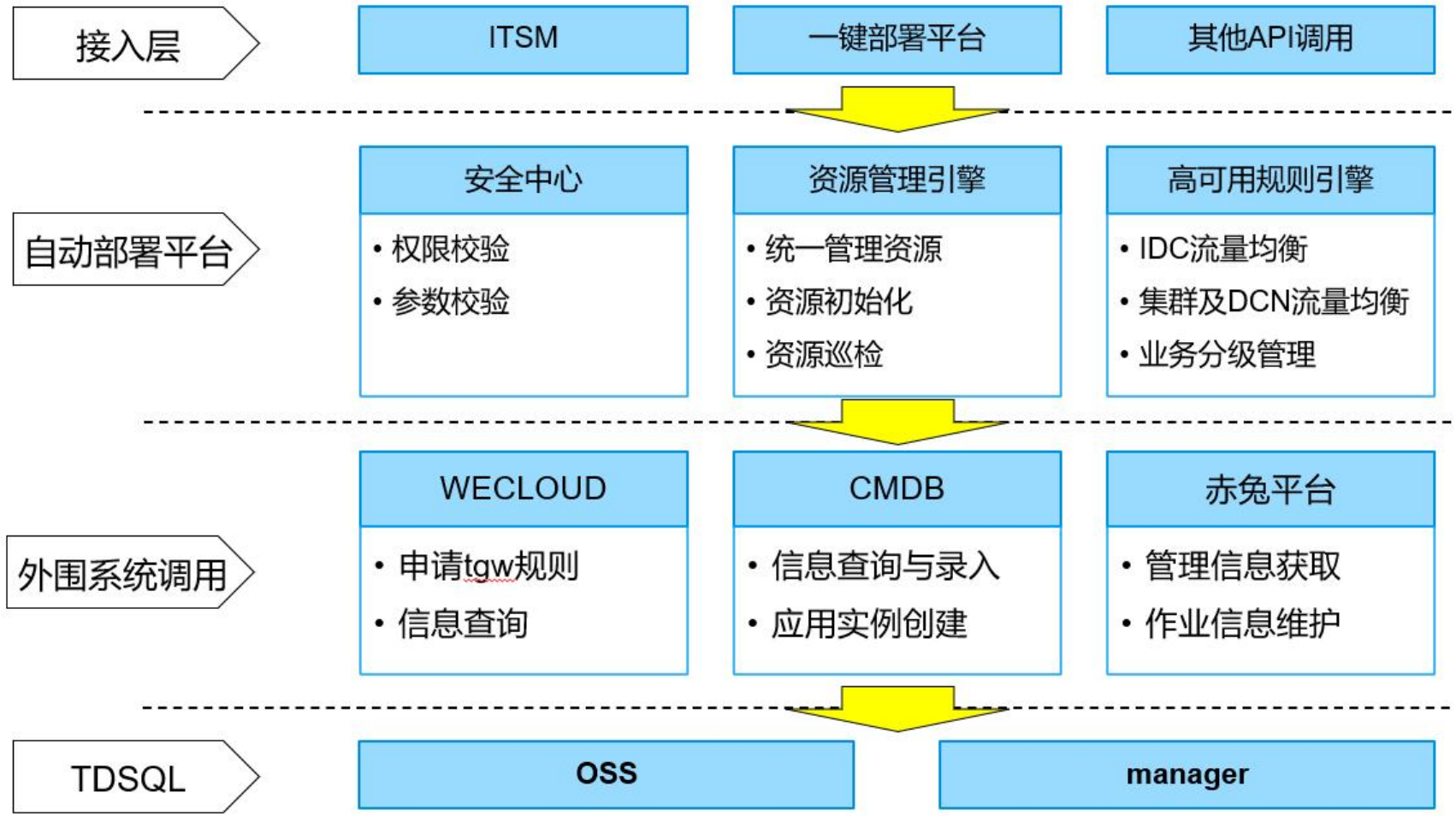
TDSQL自动巡检

巡检项	巡检目的
relay log序号超700000	relay log序号到100万后，就不会自动清理
管理节点磁盘使用率超80	保证磁盘使用率
备份情况检查	保证备份完整性
set同IDC<=2台网关存在链接	防止单点故障
无链接在线网关	防止单点故障
zk遗留任务扫描	防止主manager重启后，遗留的任务重新发起，如下线资源，主备切换等
内存使用率超90	防止OOM
FRM_FILE_OVERSIZE	分区建太多，frm大小超阈值后，表无法读写
GTID_SLAVE_POS表大小监控	MariaDB并行复制中，如果复制冲突发生回滚，插入mysql.gtid_slave_pos的数据不会随之回滚，导致该表越来越大。影响切换效率甚至超时
zookpeer备份	保证zookeeper数据备份有效
表数据一致校验	DCN间数据一致性校验
空闲机器巡检	保证空闲资源随时可用
主备DCN ZK权限一致性对比	DCN间权限一致性校验
zk域名检查	各组件配置的zookeeper列表为域命形式，保证正确
非watch节点延迟大于nS	主备复制延迟优化
OSS配置检查	OSS配置可能会影响集群监控
告警屏蔽内容列表	确认屏蔽告警的合理性
机器资源数据正确性巡检	检查集群里配置的机器资源机型、配置等信息是否正确，保证扩缩容顺利
大事务检查	检查大binlog，保证主备复制效率及数据恢复顺利

TDSQL安全管理

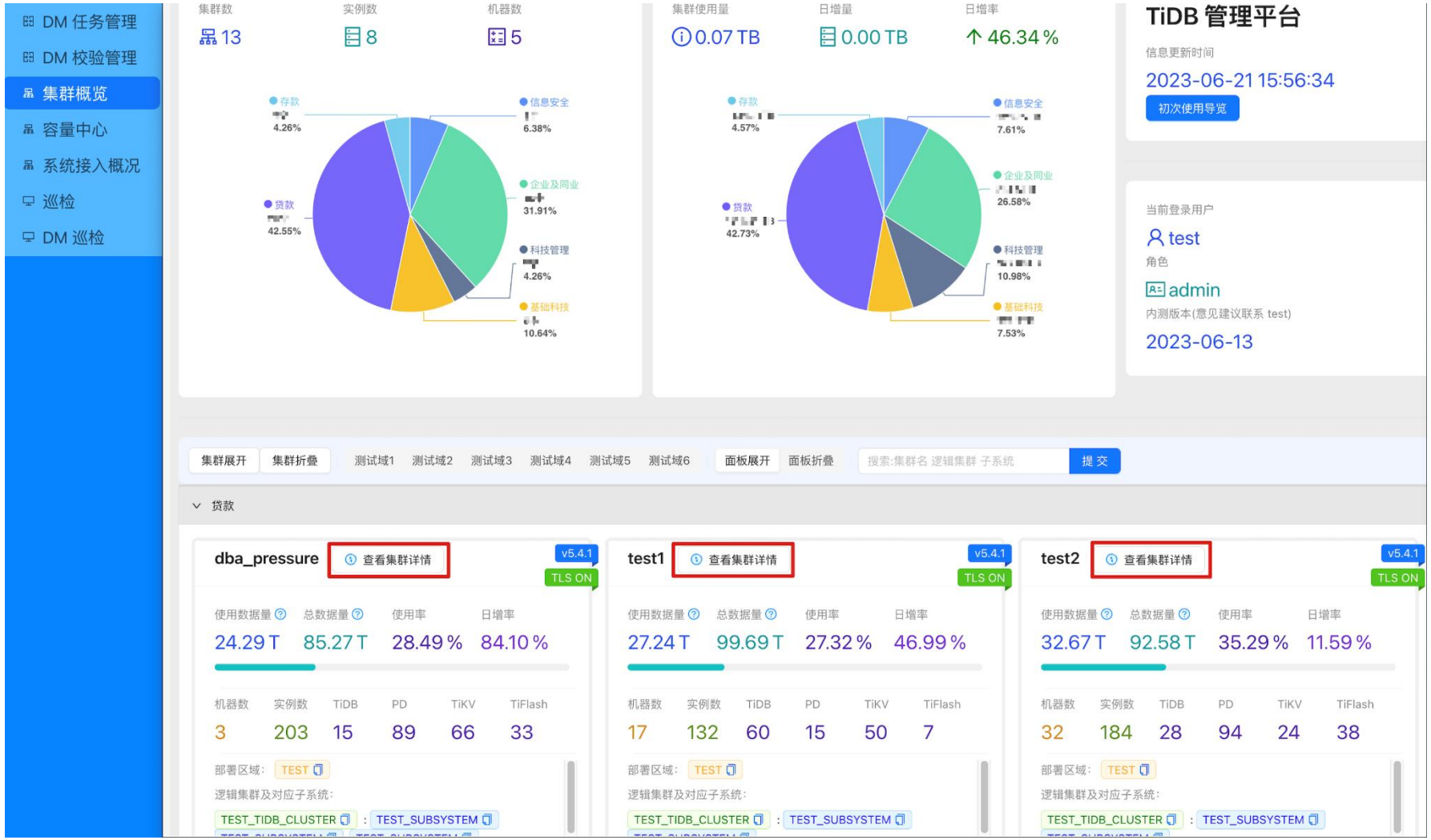

信息安全
旁路审计
proxy侧抓包
准实时预警
AOMP
自动备份数据
自动切分大事务
SQL规范检查
过滤敏感信息
DDL调用PT
防止数据误删
密码管家
生成、存储、加解密密码
提供SDK存取密码
细化用户权限
mysqlsec
过滤敏感信息
密码加密
权限优化
审计操作
覆盖mysqldump功能

TDSQL自动化一键部署

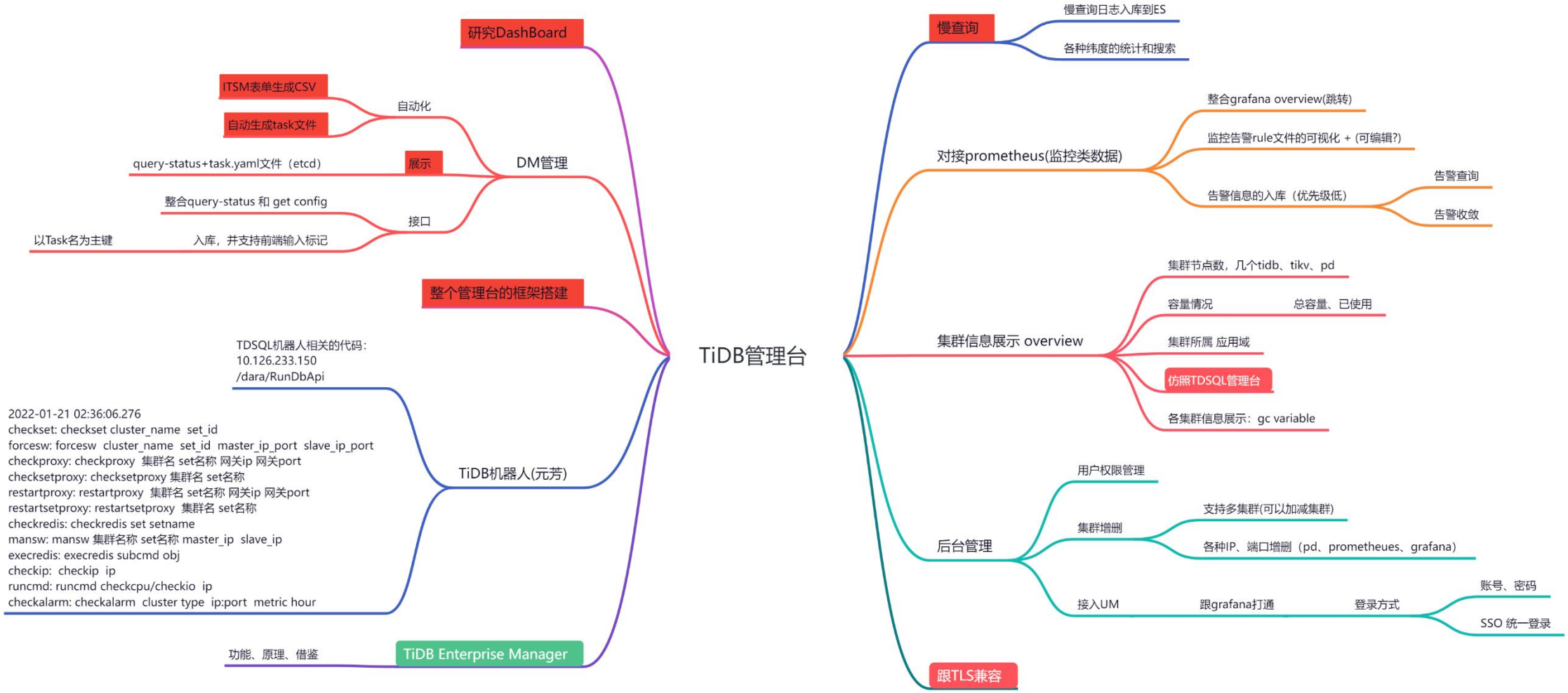


3.TiDB运维实践

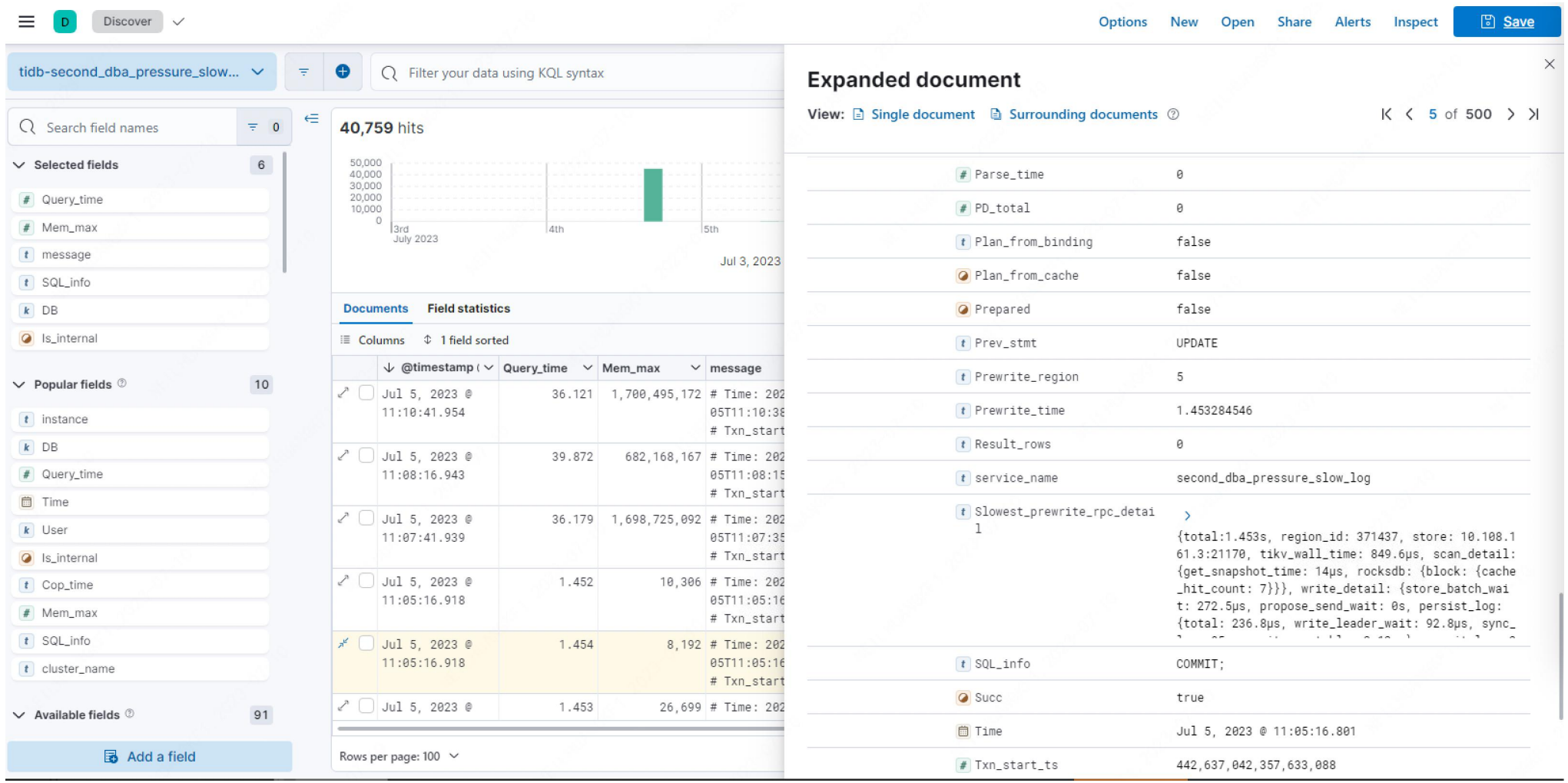
自研TiDB管理台(1/2)



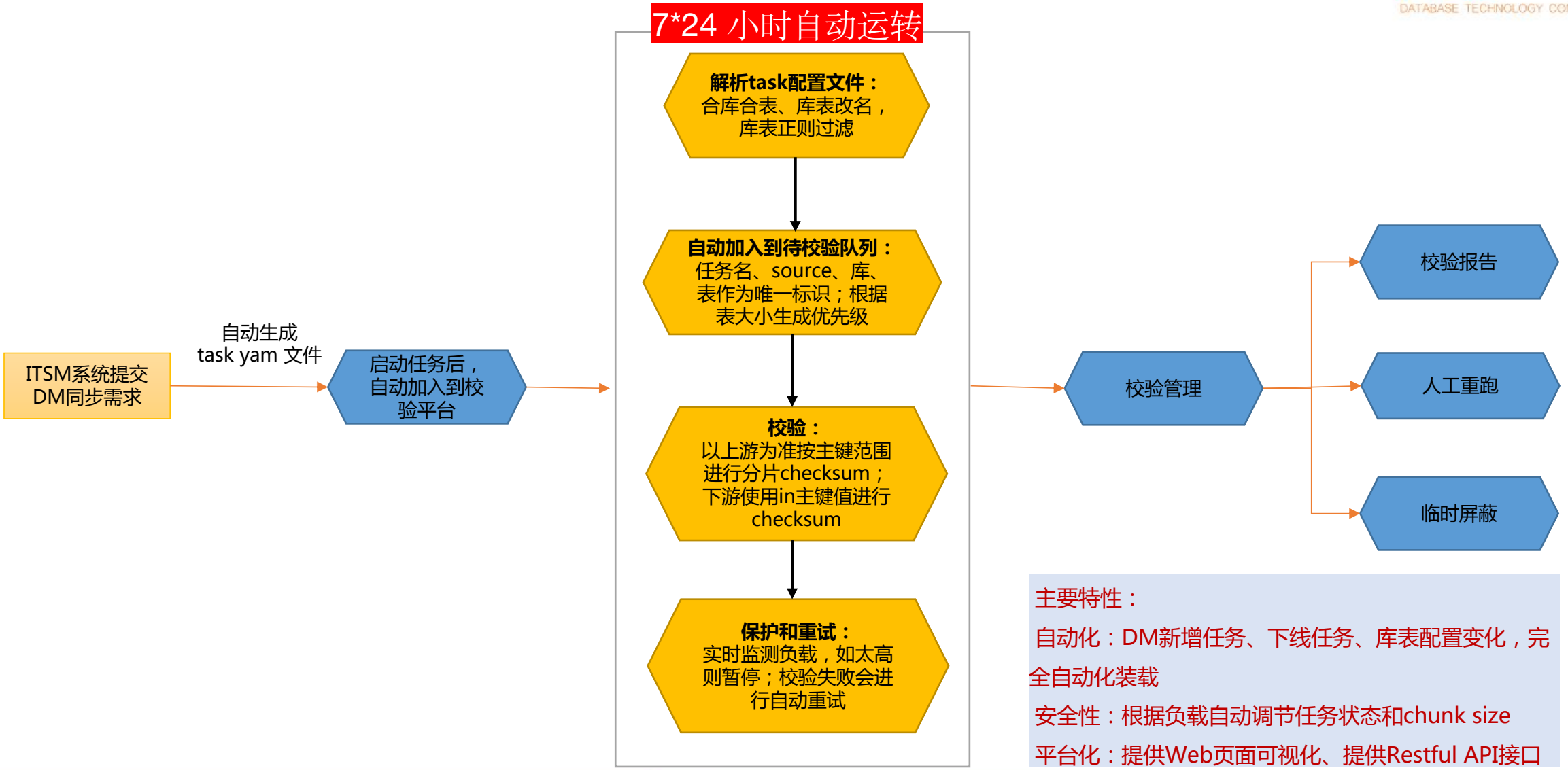
自研TiDB管理台(2/2)



基于ES的TiDB慢SQL分析平台



自研TiDB DM同步较验平台



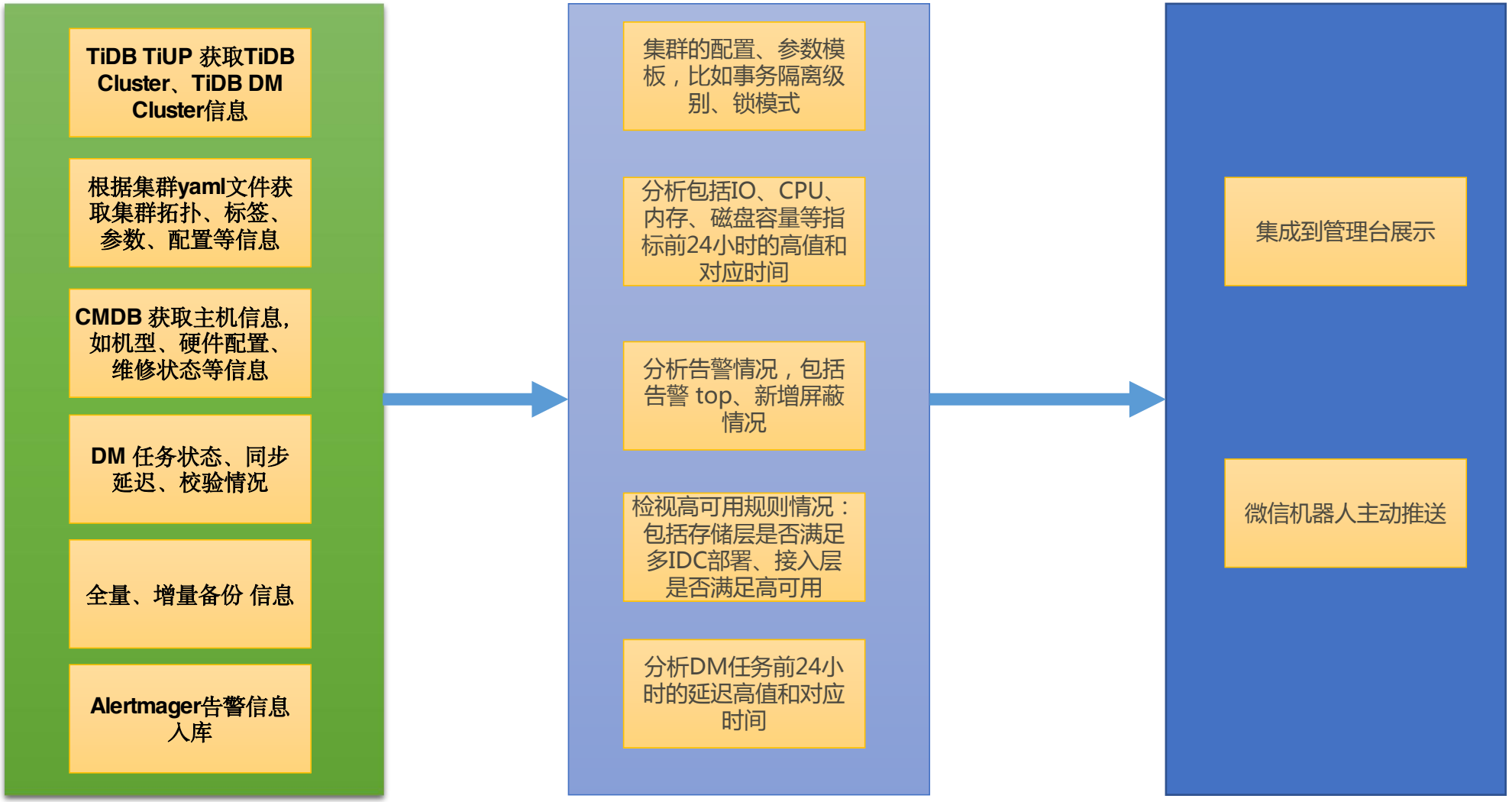
主要特性：

自动化：DM新增任务、下线任务、库表配置变化，完全自动化装载

安全性：根据负载自动调节任务状态和chunk size

平台化：提供Web页面可视化、提供Restful API接口

TiDB自动巡检


TiDB TiUP 获取TiDB Cluster、TiDB DM Cluster信息
根据集群yaml文件获取集群拓扑、标签、参数、配置等信息
CMDB 获取主机信息，如机型、硬件配置、维修状态等信息
DM 任务状态、同步延迟、校验情况
全量、增量备份 信息
Alertmager告警信息入库
集群的配置、参数模板，比如事务隔离级别、锁模式
分析包括IO、CPU、内存、磁盘容量等指标前24小时的高值和对应时间
分析告警情况，包括告警 top、新增屏蔽情况
检视高可用规则情况：包括存储层是否满足多IDC部署、接入层是否满足高可用
分析DM任务前24小时的延迟高值和对应时间
集成到管理台展示
微信机器人主动推送
数据采集层
巡检处理层
巡检展示层

THANKS