



性能优化利器 —

CDA 数据分析师
www.cda.cn

数据库审核平台实践

宜信研发中心 — 韩锋

概况简介

面临的挑战

审核平台选型

CDA 数据分析师
www.cda.cn

审核平台实践

不足及发展

概况简介

CDA 数据分析师
www.cda.cn



人人有信用 信用有价值

宜信财富
CreditEase

**如何
让财富稳健
增值**

我们不仅是固定收益理财专家
更是您可信赖的财富管理伙伴

服务大众富裕阶层，高净值人群

P2P

基金

保险

信托

私募股权

海外产品

风险投资

产品包括自主开发和第三方

宜信简介

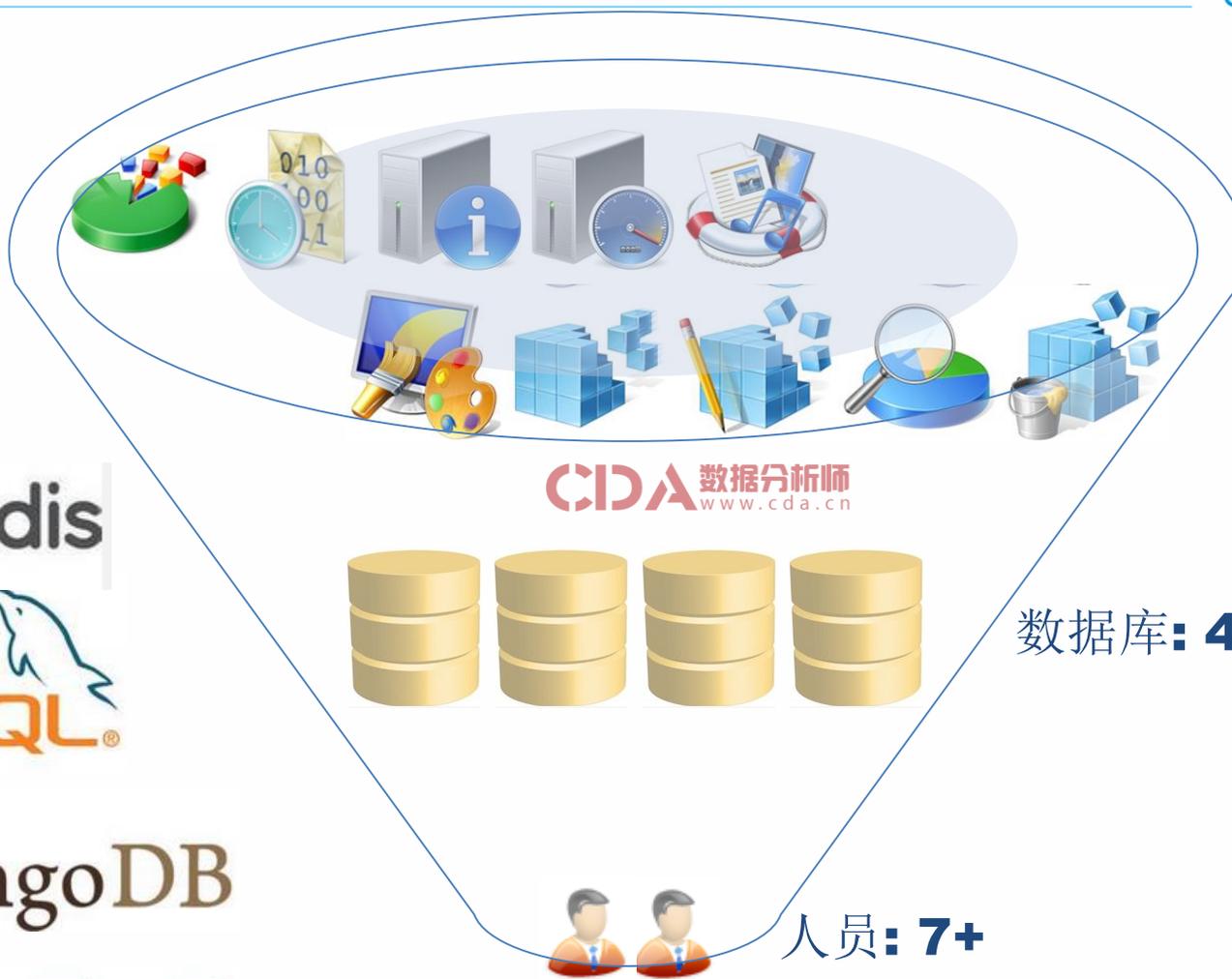


个人简介



韩锋：早年从事软件开发工作，后因个人兴趣转入数据库领域。有着多年的一线数据库架构、设计、开发经验，曾担任多家公司首席 DBA、数据库架构师等职。现就职于宜信，曾就职于当当网、TOM 在线等公司。精通多种关系型数据库，包括 Oracle、MySQL、GreenPlum、Informix 等，对 NoSQL 及大数据相关技术也很感兴趣，实践经验丰富。

运维规模及种类



系统: **200+**
开发: **1000+**

CDA 数据分析师
www.cda.cn

数据库: **40+**

人员: **7+**

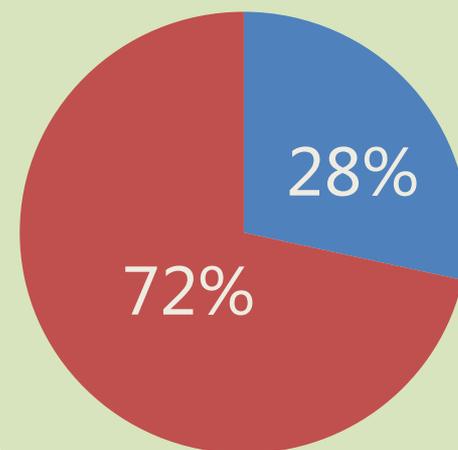


面临的挑战

CDA 数据分析师
www.cda.cn



某核心系统核心表
XXX使用SQL占比
1061 / 3730

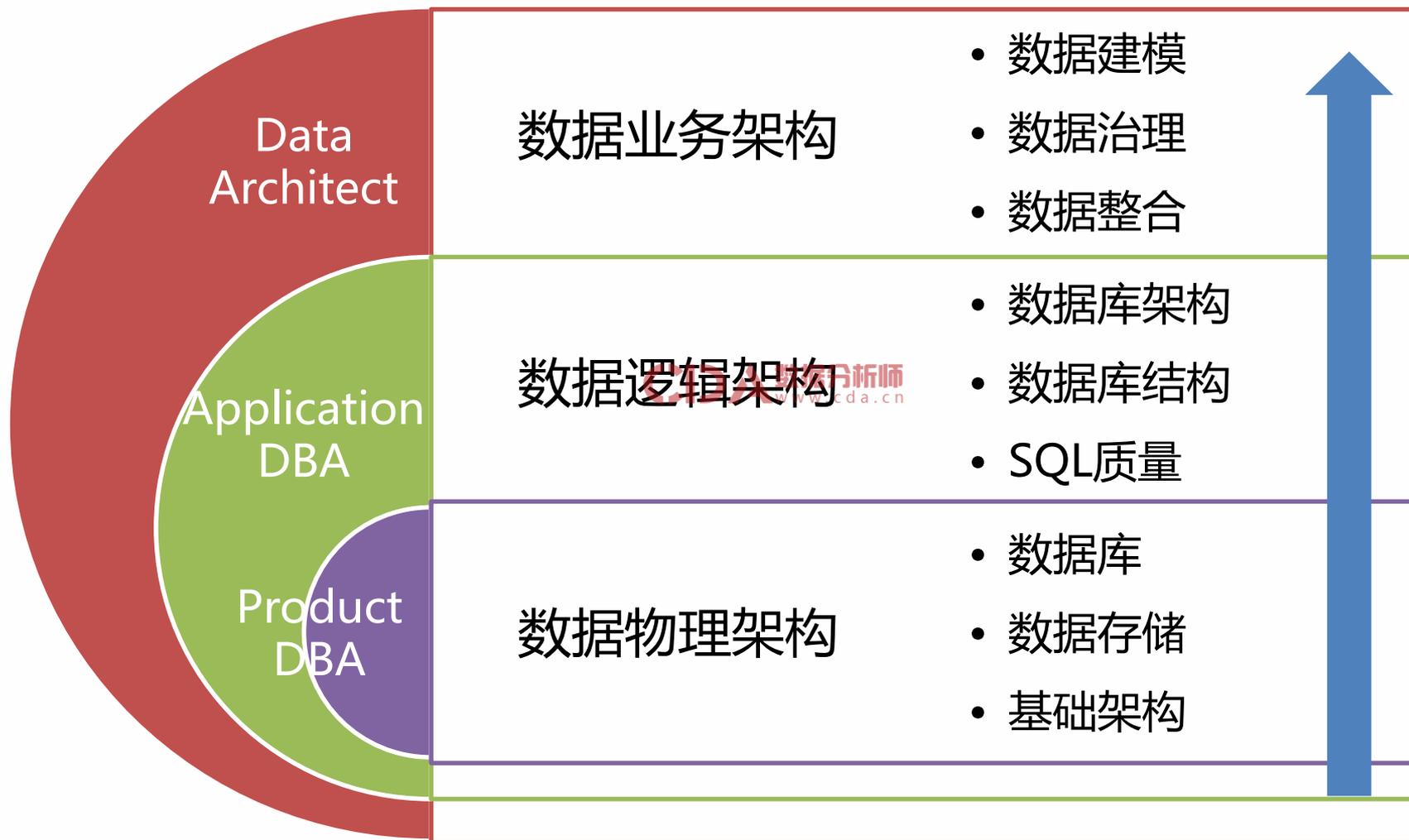


案例 – SQL语句

```
SELECT /*+ INDEX (A1 xxxxx) */ SUM(A2.CRKSL),  
      SUM(A2.CRKSL*A2.DJ) ...  
FROM xxxx A2, xxxx A1  
WHERE A2.CRKFLAG=xxx AND A2.CDATE >=xxx AND A2.CDA
```

CDA 数据分析师
www.cda.cn

Id	Operation	Name	Rows	Bytes	Cost (%CPU)	Time	Pstart	Pstop
0	SELECT STATEMENT				9890G(100)			
1	SORT AGGREGATE		1	41				
2	MERGE JOIN CARTESIAN		3505T	127P	9890G (1)	999:59:59		
3	PARTITION RANGE ITERATOR		25M	1010M	170K (1)	00:34:12	153	243
4	TABLE ACCESS FULL	XXXXXXXXXXXXXXXXXXXX	25M	1010M	170K (1)	00:34:12	153	243
5	BUFFER SORT		135M		9890G (1)	999:59:59		
6	INDEX FULL SCAN	XXXXXXXXXXXXXXXXXXXX	135M		382K (1)	01:16:34		



多种不平衡



重运维，轻架构、优化

CDA 重商业产品，轻开源产品
www.cda.cn

重手工，轻平台、工具

重初级，轻中高级

如何落地是个难题?

- Oracle结构设计规范
- Oracle开发规范
- MySQL结构设计规范
- MySQL开发规范

CDA 数据分析师
www.cda.cn



出路在哪里...



平台的选型

CDA 数据分析师
www.cda.cn

业内都是怎么做的

代表做法	特点	优点	缺点
智能分析引擎	自研SQL分析引擎，分析语句成本，并自动实现审核、分流、限流等操作。	<ul style="list-style-type: none">• 可自动审核• 扩展后可线上使用，实现分流等	<ul style="list-style-type: none">• 技术难度大• 效率不高
工具+人工审核	自研工具加后期人工审核，事后过滤、人工标记，跟踪全流程	<ul style="list-style-type: none">• 对SQL精细粒度控制，灵活度大。• 完成对SQL整个生命周期的管理。• 技术难度不大	<ul style="list-style-type: none">• 需要人工投入
商业产品	直接抽取SQL，自主分析，个别需要人工介入	<ul style="list-style-type: none">• 功能强大，有技术支持• 周期短，见效快	<ul style="list-style-type: none">• 费用较高• 扩展性差

转变思想，全民动员，人人开发

知识沉淀，做好标准，方便落地

小步快跑，落地实施，不断修正

结合自身，定制目标，学做减法

他山之石可以攻玉，大胆引进

审核平台实践

CDA 数据分析师
www.cda.cn

功能介绍



互联网从业者，致力于数据库管理，提倡工具自动化，解放自己，做更有价值的事情



Our Story



每个人都是一个传奇
每个同事都有无穷的潜力

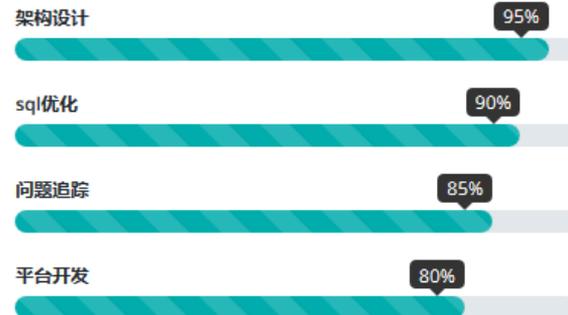
Our Philosophy

“ 我们努力工作
只是为了让自己少做工
作 ”



Sean Ngu
Front End Developer

我们的经验



背景介绍 — 大平台功能

仪表盘

业务报表 性能报表

结构元数据

查询 同步 导出

数据库审核

结构 SQL

平台元数据

查询 API

服务托管

..

数据变更

性能管理

数据管理 分析报告 慢查询

资源管理

部署 变更 回收 迁移 扩容

数据管理

备份 导入导出 校验 同步

健康管理

巡检 故障 诊断

对象管理

健康 空间 访问 性能

数据安全

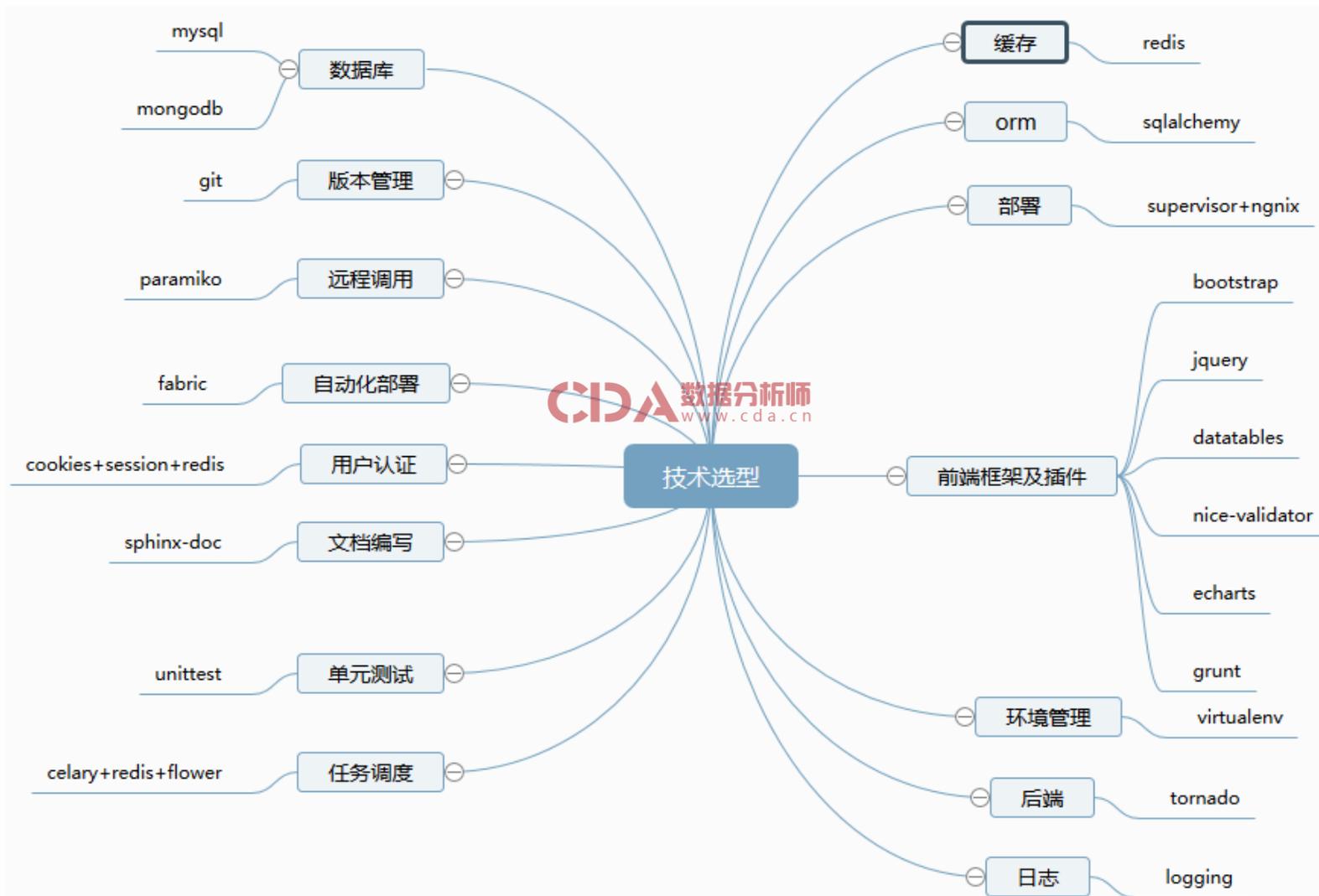
脱敏 审计 策略 授权

平台管理

用户 权限 日志 作业

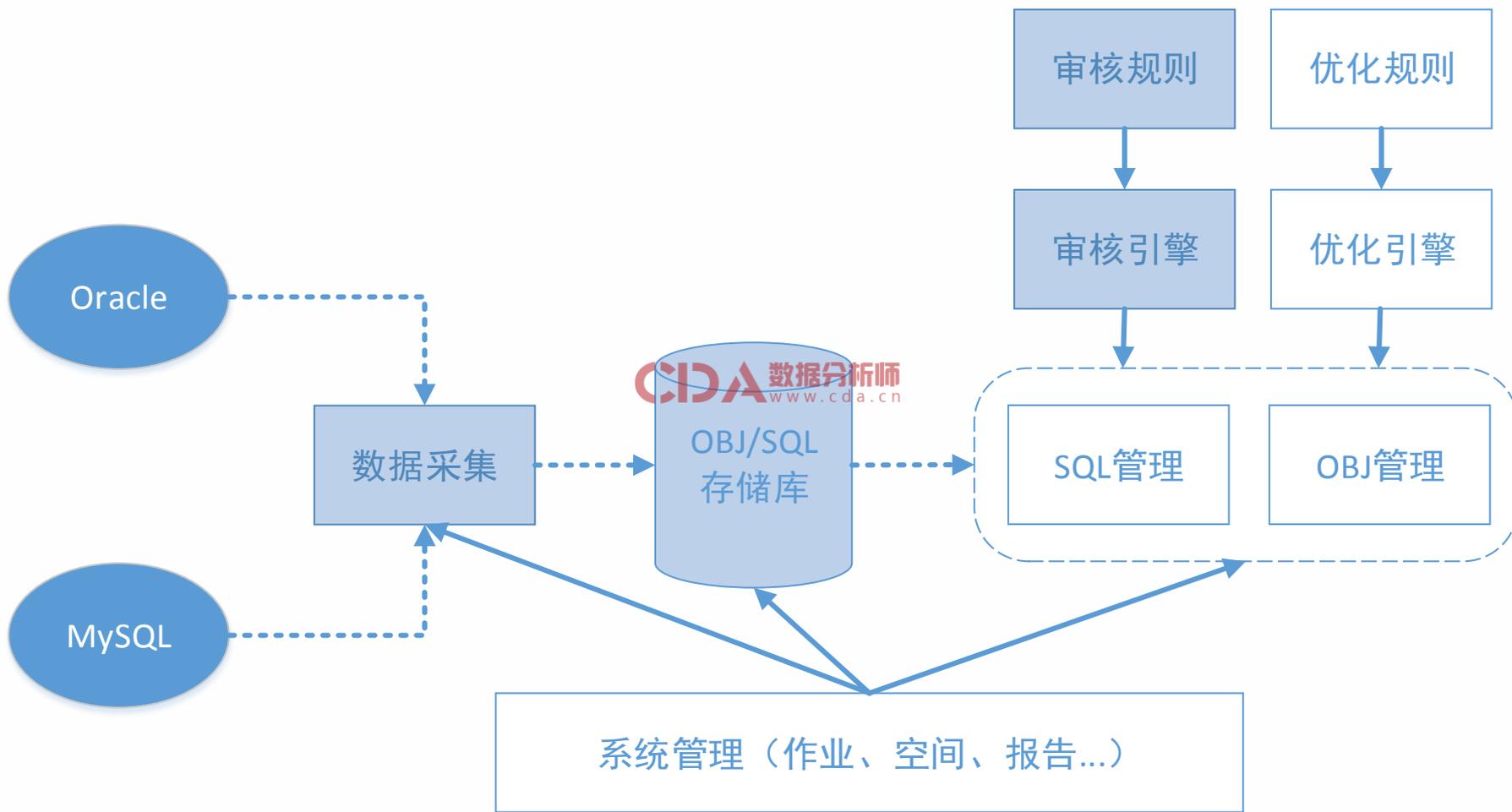
12个子系统，36个模块

背景介绍 — 技术栈

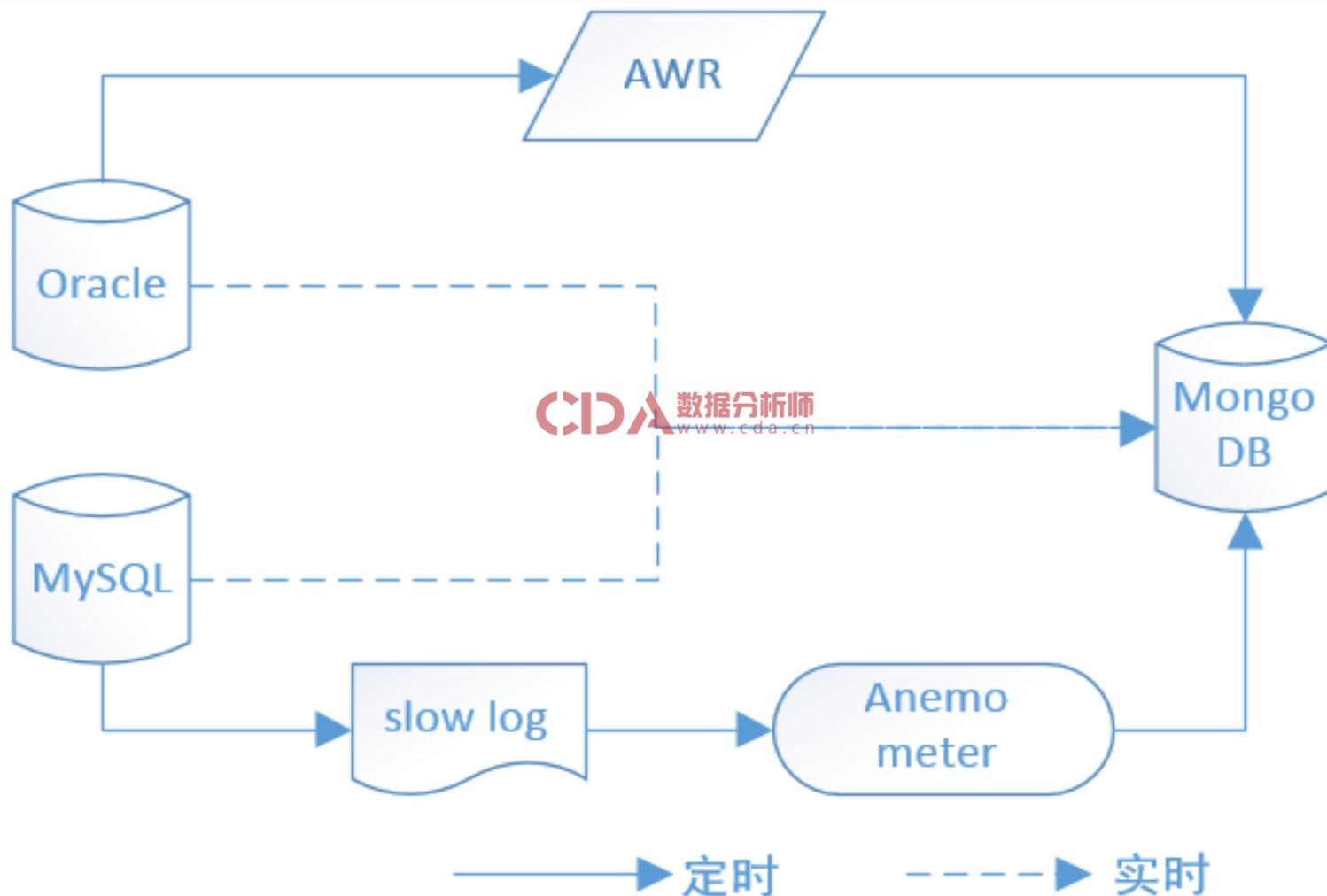


审核类别	示例规则
对象级	大表未分区
	未创建主键
语句级	多表关联
	标量子查询
执行计划级	大表全表扫描
	笛卡尔积
执行特征级	扫描块数与返回记录比例过低
	子游标数过多

平台设计 — 架构设计简图



	抓取项	Oracle	MySQL
对象	统计信息	√	√*
	存储特征	√	√*
	结构信息	√	
	访问特征	√	
SQL	SQL文本	√	√
	执行计划	√	
	游标	√	
	绑定变量	√	
	执行特征	√	√



平台设计 — 信息存储格式

```
mysql> explain format=json select * from film where film_id in
***** 1. row *****
EXPLAIN: {
  "query_block": {
    "select_id": 1,
    "nested_loop": [
      {
        "table": {
          "table_name": "film_actor",
          "access_type": "ref",
          "possible_keys": [
            "PRIMARY",
            "idx_fk_film_id"
          ],
          "key": "PRIMARY",
          "used_key_parts": [
            "actor_id"
          ],
          "key_length": "2",
          "ref": [
            "const"
          ],
          "rows": 19,
          "filtered": 100,
          "using_index": true
        }
      },
      {
        "table": {
          "table_name": "film",
          "access_type": "const",
          "possible_keys": [
            "PRIMARY"
          ],
          "key": "PRIMARY",
          "used_key_parts": [
            "film_id"
          ],
          "key_length": "4",
          "ref": [
            "const"
          ],
          "rows": 1,
          "filtered": 100,
          "using_index": true
        }
      }
    ]
  }
}
```

schemaless



平台功能 — 规则定义（对象级）

规则类别	规则说明
表、分区	大表过多
	超多指定规模没有分区
	单表或单分区数据量过多
	存在并行属性
索引	分区数量过多
	外键没有索引
	字段重复索引
	聚簇因子多大索引
字段	字段重复索引
	字段数量过多
	记录长度多长
	字段类型不匹配

平台功能 — 规则定义（执行计划级）

规则类别	规则说明
访问路径	大表扫描
	大索引扫描
	大索引快速全扫描
	索引跳跃扫描
	分区全扫描
	非连续分区扫描
	跨分区扫描
表间关联	笛卡尔积
	多表关联
	嵌套循环层次过多
类型转换	存在隐式类型转换
绑定变量	未使用绑定变量

平台功能 — 规则定义（执行特征级）

规则说明

扫描块数与返回记录数比例过低

子游标过多

elapsed_time

cpu_time

buffer_gets

CDA 数据分析师
www.cda.cn

disk_reads

direct_writes

Executions

平台功能 — 规则定义（语句级）

规则说明

select *

嵌套select子句

谓词中出现反向操作符

多个过滤条件通过or连接

存在子查询

CDA 数据分析师
www.cda.cn

存在三个以上的多表关联

存在全连接或外连接

delete中必须出现where

update中出现order by子句

inlist元素过多

重复查询子句

出现union集合操作

大表全表扫描规则 →

Id	Operation	Name	Rows	Bytes	Cost (%CPU)	Time
0	SELECT STATEMENT		1	8	410 (1)	00:00:05
* 1	TABLE ACCESS FULL	T_TEST	1	8	410 (1)	00:00:05

```
db.sqlplan.find({'SQL_ID':'fzzxntg1g9u6r'},  
  {'OBJECT_TYPE:1,OBJECT_NAME:1,OPERATION_DISPLAY:1,OPTIONS:1,_id:0})
```

```
{ "O" : "SELECT STATEMENT", "OPTIONS" : null, "OBJECT_NAME" : null }
```

```
{ "O" : " TABLE ACCESS", "OPTIONS" : "FULL", "OBJECT_NAME" : "T_TEST" }
```

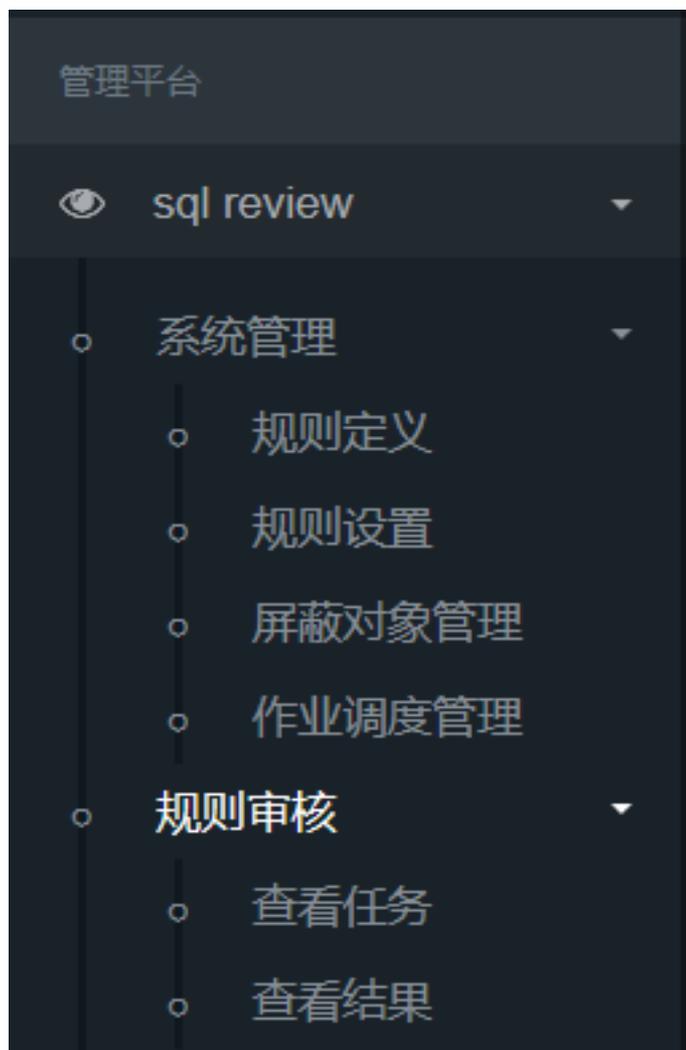
平台实现 — 规则实现示例

```
db.@sql@.find({"OPERATION":"TABLE ACCESS",
              "OPTIONS":"FULL",
              "USERNAME":"@username@",
              "ETL_DATE":"@etl_date@"}
).forEach(function(x)
{
  if(db.obj_tab_info.findOne({
    "TABLE_NAME":x.OBJECT_NAME,
    $or:
    [{"NUM_ROWS":{"$gt:@params1@"}},
     {"PHY_SIZE(MB)":{"$gt:@params0@"}}
    ]})
  )db.@tmp@.save({
    "SQL_ID":x.SQL_ID,
    "PLAN_HASH_VALUE":x.PLAN_HASH_VALUE,
    "OBJECT_NAME":x.OBJECT_NAME});});"
```

CDA 数据分析师
www.cda.cn

- 
- 过滤执行计划
 - 按统计信息筛选
 - 按存储特征筛选
 - 获得执行计划
 - 后台存储并计分
 - 汇总展示报告

平台界面 — 主菜单



CDA 数据分析师
www.cda.cn

平台界面—制定审核任务

ip地址：	<input type="text" value="10.100.30.153"/>			
实例端口：	<input type="text"/>			
schema：	<input type="text" value="schema"/>			
日期选择：	<input type="text" value="开始日期"/> <input type="text" value="到"/> <input type="text" value="结束日期"/>			
规则类型：	<input type="checkbox"/> plan 特征	<input type="checkbox"/> text	<input type="checkbox"/> 结构	<input type="checkbox"/> 执行
	<input type="button" value="执行任务"/>			

CDA 数据分析师
www.cda.cn

平台界面— 查看审核任务结果

导出报告

查看报告

操作用户	用户名	创建时间	类型	状态	选择
system	username	2016-08-18 18:43:04	成功	obj	<input type="checkbox"/>
system	username	2016-08-18 18:43:34	成功	obj	<input type="checkbox"/>
system	username	2016-08-18 18:43:54	成功	obj	<input type="checkbox"/>
system	username	2016-08-18 18:43:56	成功	obj	<input type="checkbox"/>
system	username	2016-08-18 18:44:19	成功	obj	<input type="checkbox"/>
system	username	2016-08-18 18:44:21	成功	obj	<input type="checkbox"/>
system	username	2016-08-18 18:44:43	成功	obj	<input type="checkbox"/>
system	username	2016-08-18 18:44:46	成功	obj	<input type="checkbox"/>

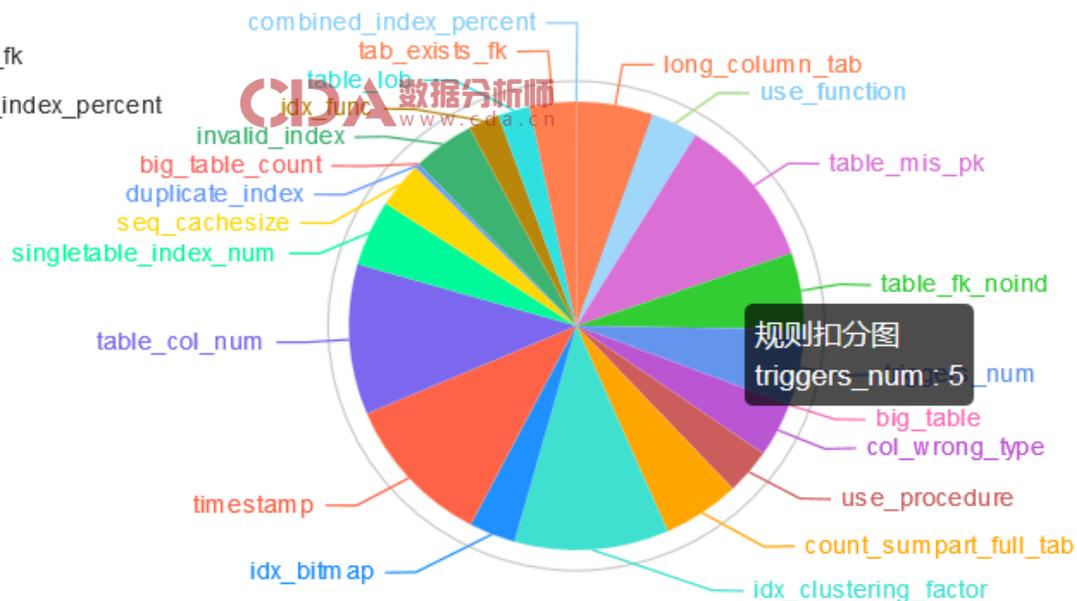
平台界面—对象审核结果概览

规则扣分

- long_column_tab
- use_function
- table_mis_pk
- table_fk_noind
- triggers_num
- big_table
- col_wrong_type
- use_procedure
- count_sumpart_full_tab
- idx_clustering_factor
- idx_bitmap
- timestamp
- table_col_num
- singletable_index_num
- seq_cachesize
- duplicate_index
- big_table_count
- invalid_index
- idx_func
- table_lob
- tab_exists_fk
- combined_index_percent

规则扣分

规则扣分详情



平台界面—对象审核结果明细

规则名称	规则描述	参数个数	违反次数	扣分
big_table	超过指定规模且没有分区的表	1	0	0
big_table_count	大表数量	1	1	0
col_wrong_type	表字段类型不匹配	3	153	3.558139534883721
combined_index_percent	组合索引数量过多或没有索引	1	13	0
count_sumpart_full_tab	分区表数量过多	1	11	5
duplicate_index	字段重复索引	0	1	0.3
idx_bitmap	是否使用位图索引	0	6	3
idx_clustering_factor	索引的聚簇因子	1	238	10

短期规划



- 完善现有规则
- 丰富前端展示
- 统一管理配置

中期规划



- 新增优化规则
- SQL全生命周期管理

长期规划



- 纳入其他数据库
- 条件成熟，考虑开源

个人作品及联系方式



谢谢！
CDA 数据分析师
www.cda.cn