



Beyebe

# 中信卡中心灵活查询平台建设方案

蔡二丰 2018-08

# 内容提要



- 1. 公司介绍

- 2. 需求理解

- 3. 建设思路

- 4. 解决方案

- 5. 性能评测

- 6. 实施方案

# 公司介绍

1. 基本介绍 2. 核心技术 3. 竞争优势 4. 成功案例

# 深圳市比一比网络科技有限公司

2009

## 公司成立

- 致力于大数据技术开发和服务，为客户提供全方位大数据解决方案

30 多项

## 软件著作权

- 数据抓取
- 数据处理
- 数据分析
- 数据搜索
- 数据可视化

9 年

## 大数据积累

- Base底层核心
- 卷积神经算法应用(图片识别)
- 隐马尔科夫模型应用(舆情)

4 项

## 发明专利

- 智能推荐
- 数据抽取
- 图像识别
- 混合索引

ISO  
ISO9001  
认证企业

CMMI  
CMMI3认  
证软件企业

高新  
国家高新  
技术企业

副会长  
深互联网技  
术应用协会

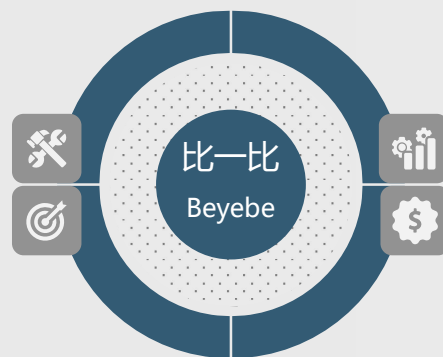
# 我们的核心技术

## 数据采集

- 数据抽取
- 图像识别
- 变身术
- 中英文混合搜索

## 数据挖掘

- 聚类
- 分析
- 关联
- 预测



## 计算引擎

- 分布式计算
- 内存计算
- 流计算
- 图计算

## 人工智能

- 自然语言处理
- 人工智能模块
- 卷积神经网络
- 机器学习算法库

# 我们的竞争优势

01

## 分布式数据整合能力

- 外部大数据收集
- 全渠道数据融合
- 数据价值提升

02

## 搜索引擎技术经验

- 多年的搜索引擎技术经验
- 对大量数据的索引建立有丰富的经验

03

## 自有发明专利技术

- 大数据云采集
- 分布式存储Bbase
- 数据抽取
- 验证码识别
- 混合索引

04

## AI智能训练和应用

- 卷积神经算法 (图片识别)
- 随机深林算法 (自动打分)
- 贝叶斯模型 (水果店筛选)
- 隐马尔科夫模型 (负面舆情消息)
- 支持向量回归 (销量预测)

2009 购物搜索引擎

2012 混合搜索引擎

2014 数据采集分析

2016 大数据产品

2017 行业应用方案

# 成功案例



## 金道集团 财经资讯采集

客户自建的采集系统从网络资讯发布到完成采集并展示需要30秒，比一比方案仅需3秒，性能整整提升10倍

## 深圳农商 信贷风控数据管理平台

采集银行授信信息补充外围工商司法等信息，数据清洗、数据分析、建立风控信贷图谱，为信贷业务提供风控辅助决策

## 广州旅游局 智慧旅游大数据平台

建立旅游数据平台系统，形成以旅游大数据为核心的智慧旅游公共服务体系，提高旅游数据在旅游业发展的应用

## 银达担保 大数据智能风控

多渠道采集全量信息，包括完整企业画像，项目招投标信息，企业舆情等，清洗的可视化呈现并且实时监控追踪

## 百果园 大数据运营决策平台

全网全屏数据监控，采集包括美团外卖、饿了么、百度外卖等O2O网站信息，进行大数据建模分析为运营决策提供数据依据，优化品牌分布、提高市场占有率

## 其他

浙江移动项目，数据收集功能由我们提供  
贵阳大数据项目，请我们去数据收集模块  
.....

# 解决方案

1. 需求理解 2. 建设思路 3. 解决方案 4. 性能评测

## 中信卡中心灵活查询平台建设



- 建设中信卡中心灵活查询平台，提供页面查询和API对接查询，满足客服或业务人员以及内部系统对卡中心各类业务数据包括历史数据的快速灵活查询需求。

- ✓ 大数据量整合
- ✓ 高并发请求
- ✓ 灵活复杂查询

### 查询引擎

- 多数据源接入
- 高性能计算
- 性能监控

### 数据API

- 路由控制
- 安全检测
- 参数校验

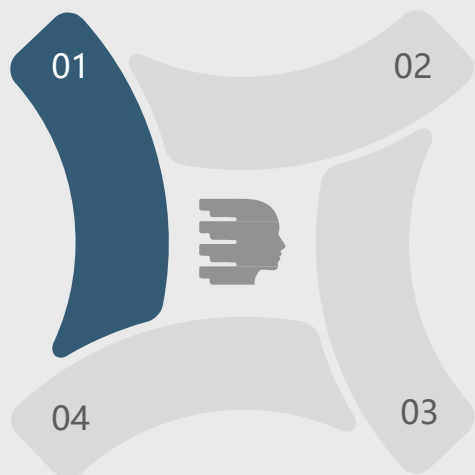
### 查询平台

- 多方式查询
- 灵活可配置
- 模板化

### 系统管理

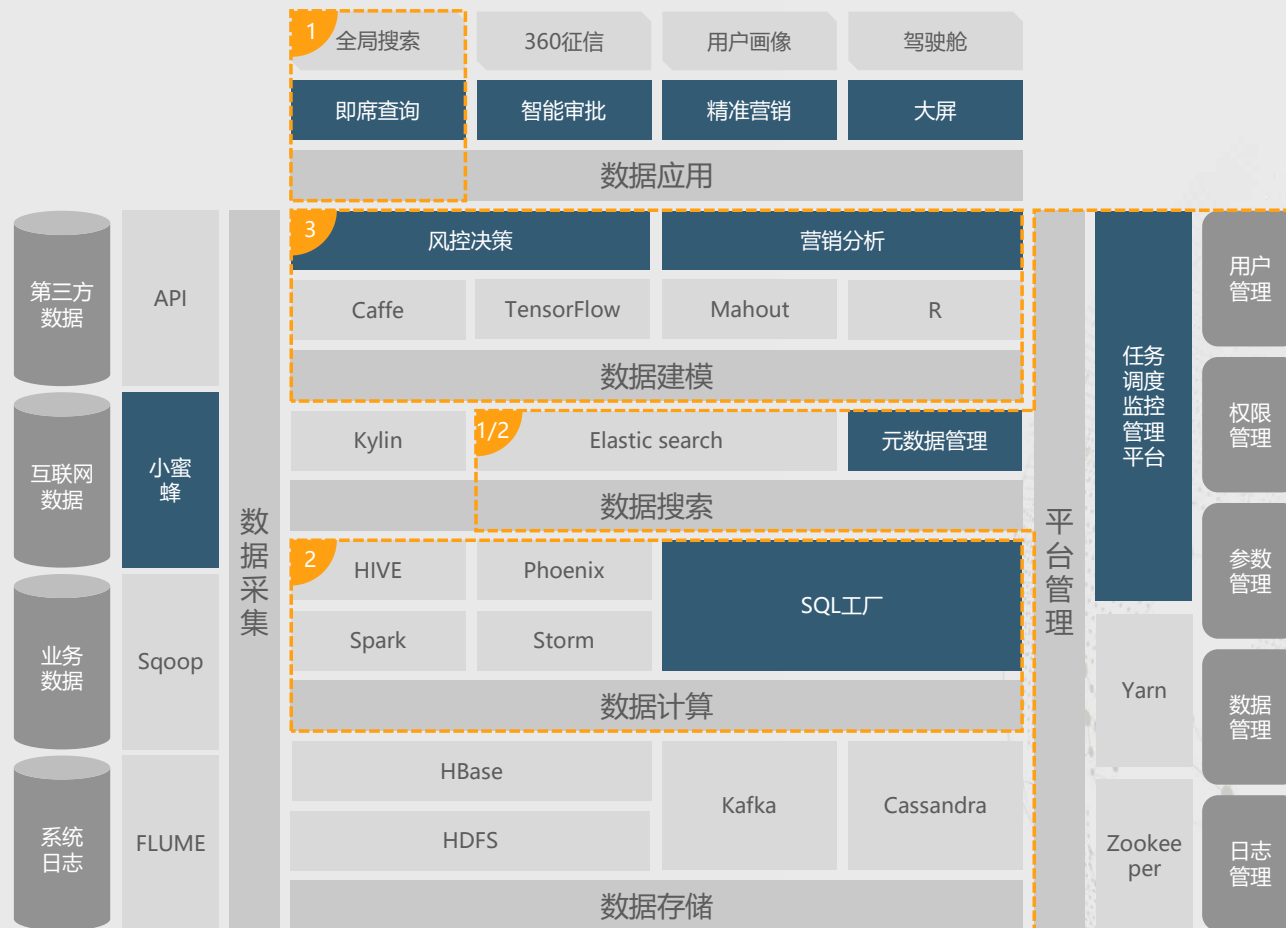
- 数据管理
- 安全管理
- 运维管理

# 建设思路

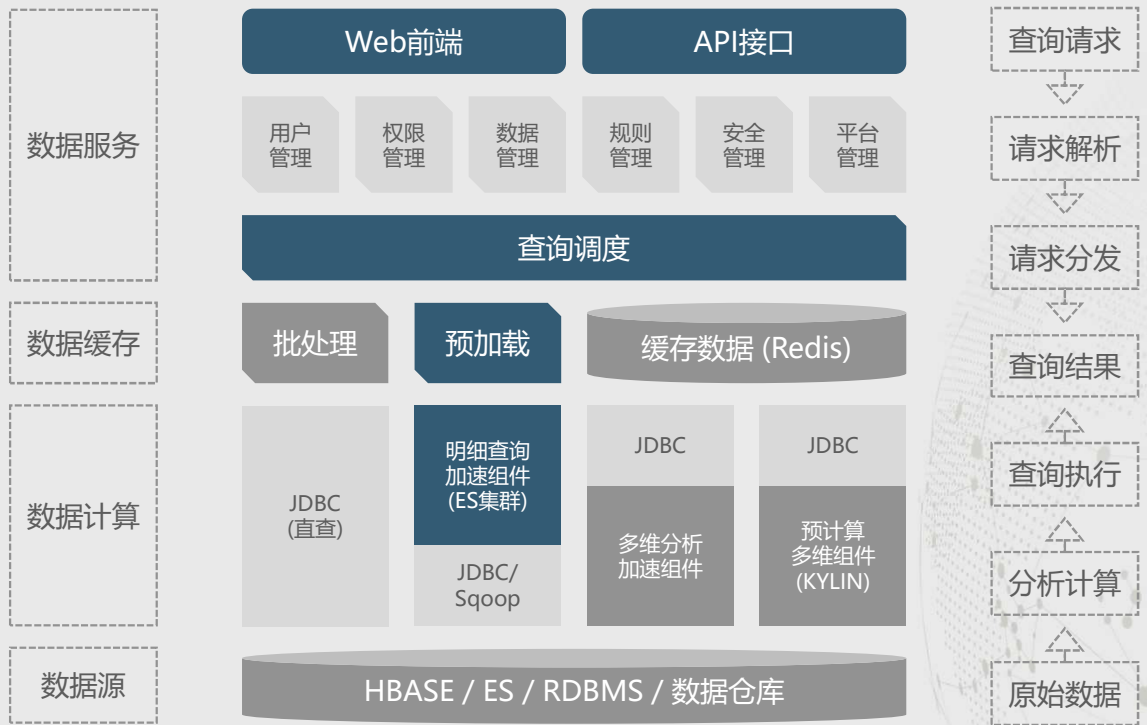


- ✓ **高性能高性价比**  
优先选择通过软件技术手段和优化系统设计实现提升查询引擎整体性能，尽量避免通过硬件堆砌提升性能的高硬件成本方案
- ✓ **高可靠高可扩展性**  
提供7\*24小时服务能力，尽量减少系统运行过程中需要重启或停止服务的场景，大量采用参数化设计，通过配置驱动系统运行
- ✓ **灵活丰富的查询能力**  
深刻理解和挖掘灵活查询平台用户使用场景，提供尽可能灵活的查询功能，提升用户体验减少二次定制化开发的需要
- ✓ **安全全面的系统管理能力**  
提供包括用户管理、权限管理、性能监控、日志管理等安全管理功能，实时监控系统运行情况，快速捕捉系统异常及时定位问题解决

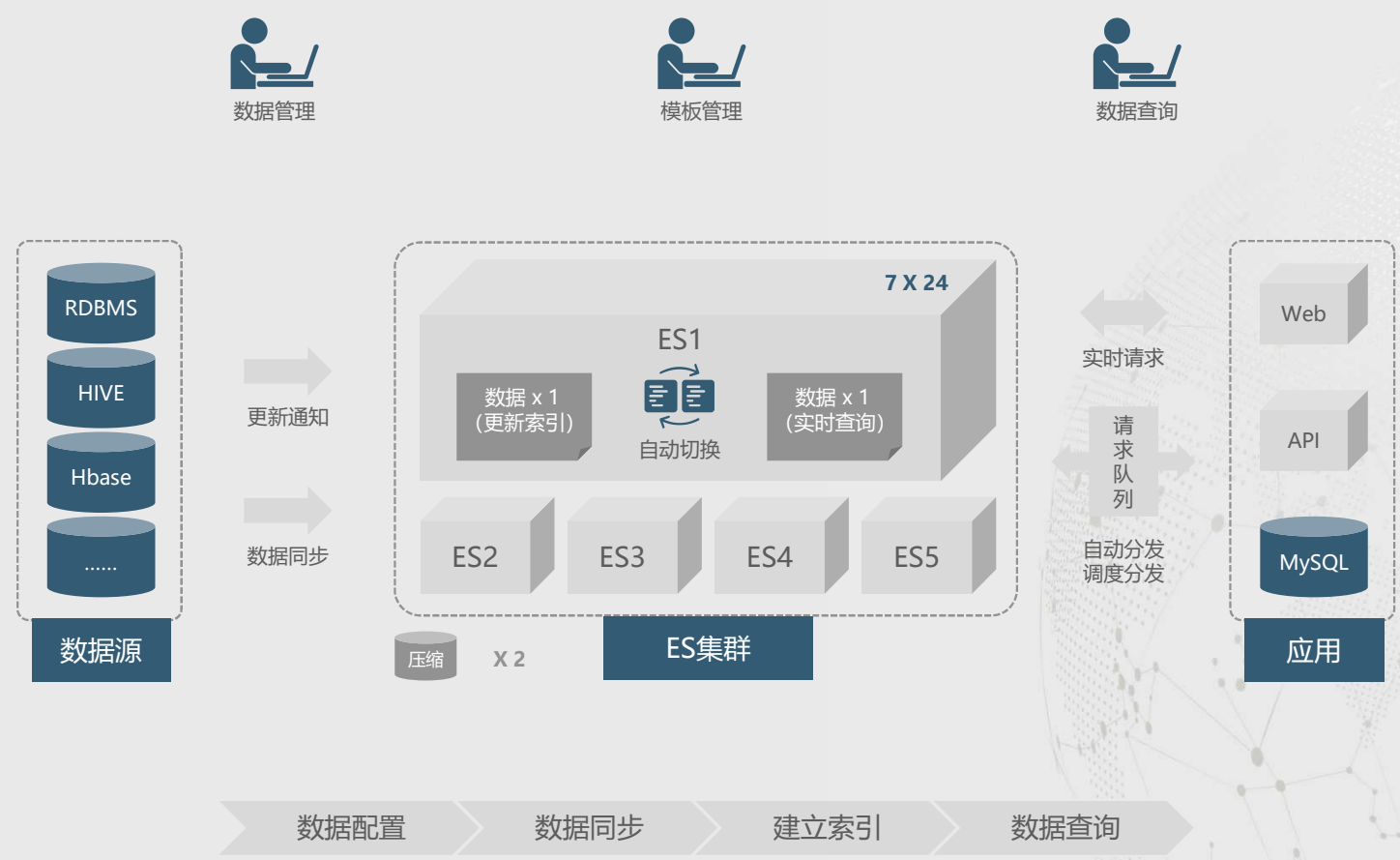
# 平台定位



# 系统方案



# 系统方案

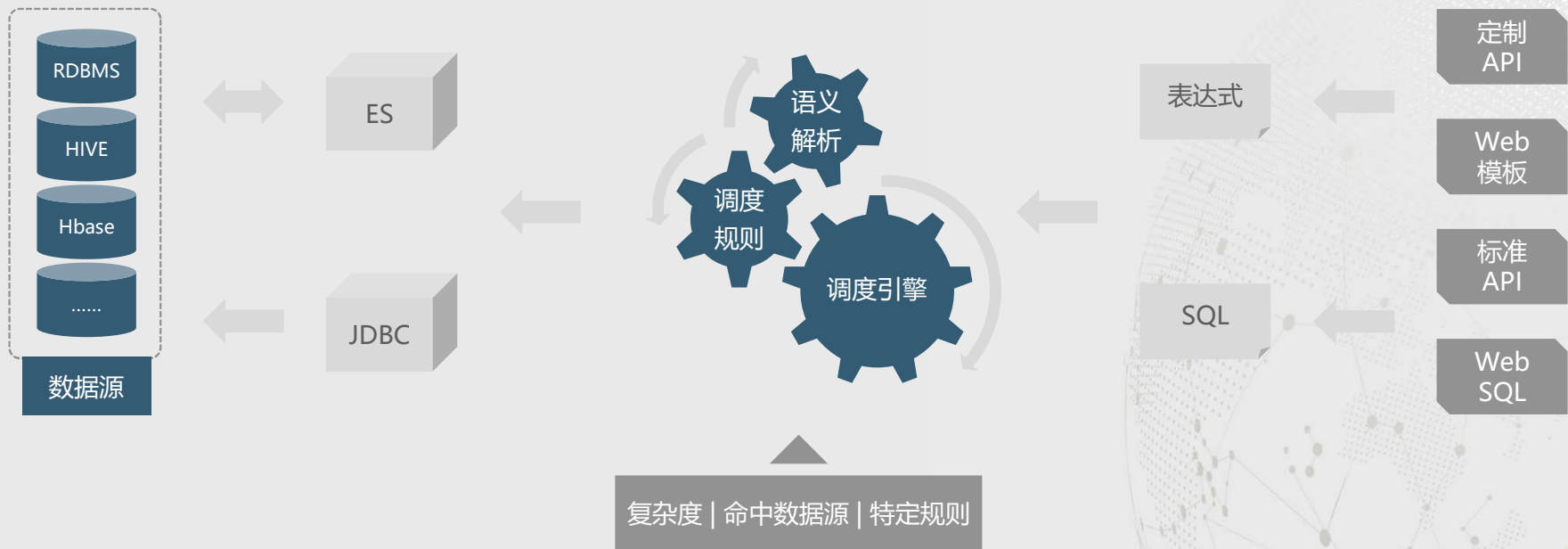


# 范围说明

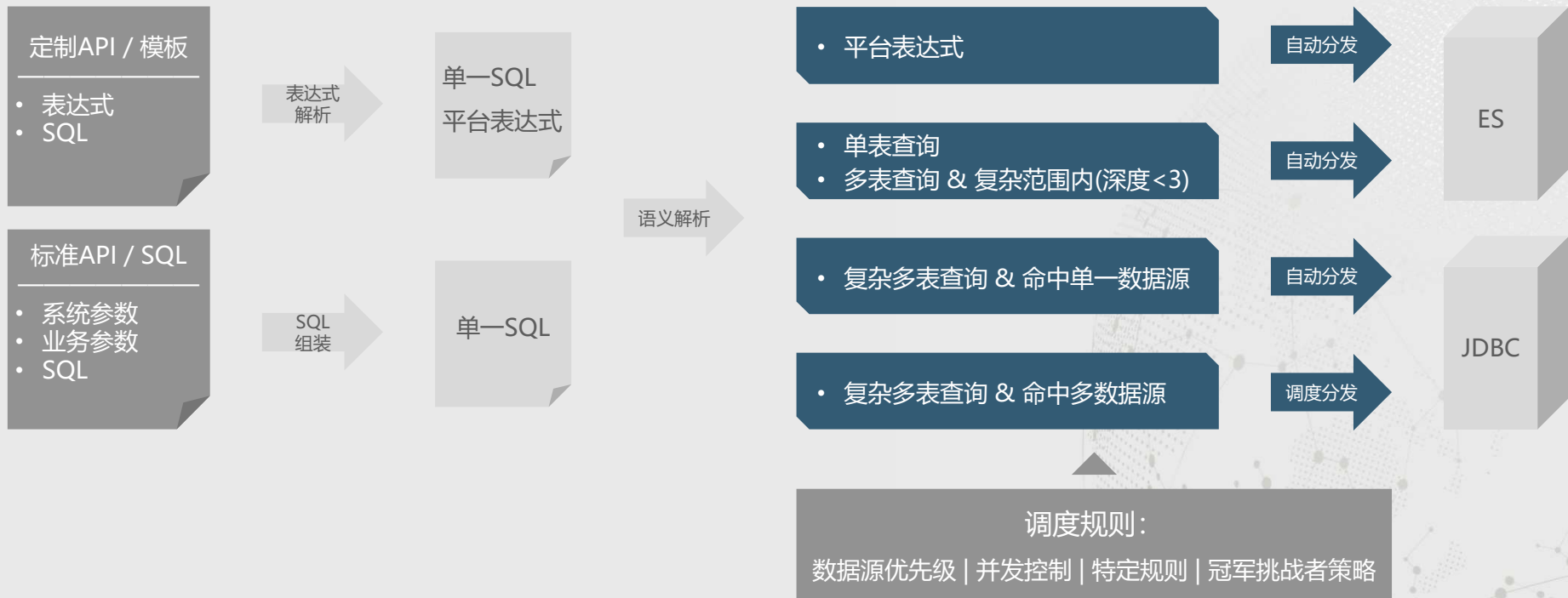
实现范围	单表	关联
数据量查询count(1)	√	/
单条件精确匹配	√	√
单条件模糊匹配	√	√
多条件精确匹配	√	√
多条件模糊匹配	√	√
排序	√	√
单字段group by	√	/
多字段group by	√	/
复杂SQL	√	√
分页查询, 每页20条	√	√

- 平台定位为面向详情的检索查询,不负责 分析、统计、分组、BI等
- 技术上复杂的统计和分组聚合操作可通过索引实现, 但考虑数据量和资源情况, 不推荐使用
- 未来如果支持更复杂的统计分析需增加大数据计算处理技术

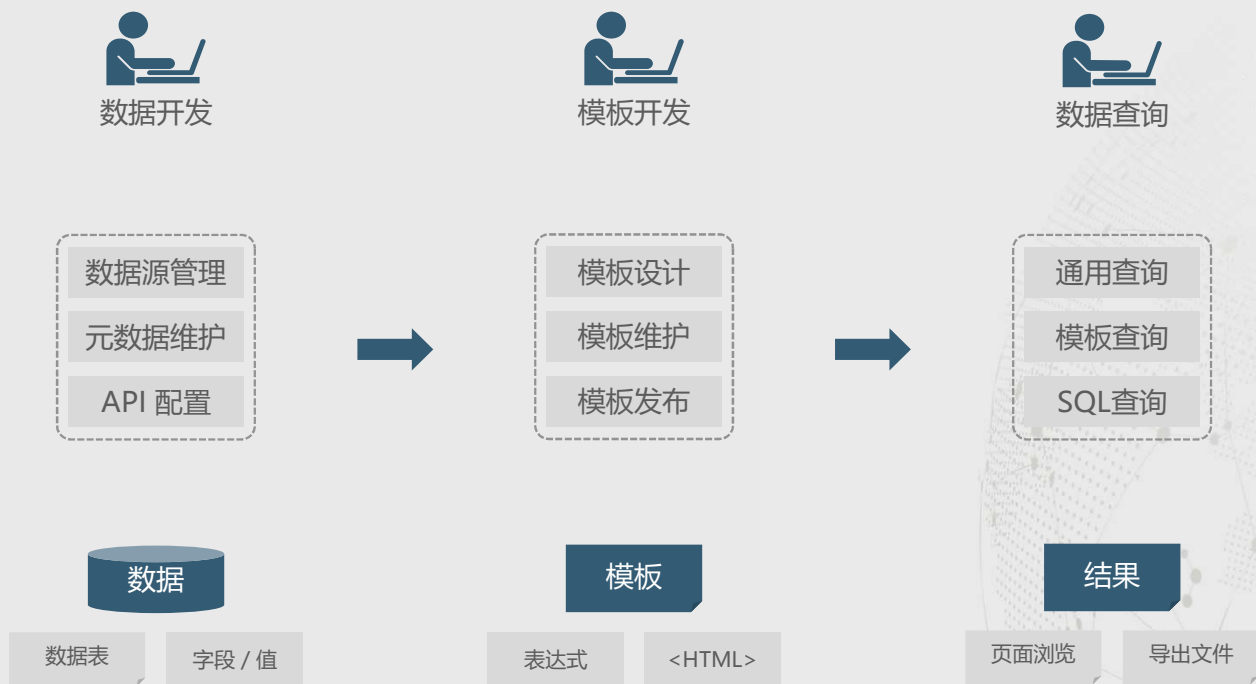
# 调度方案



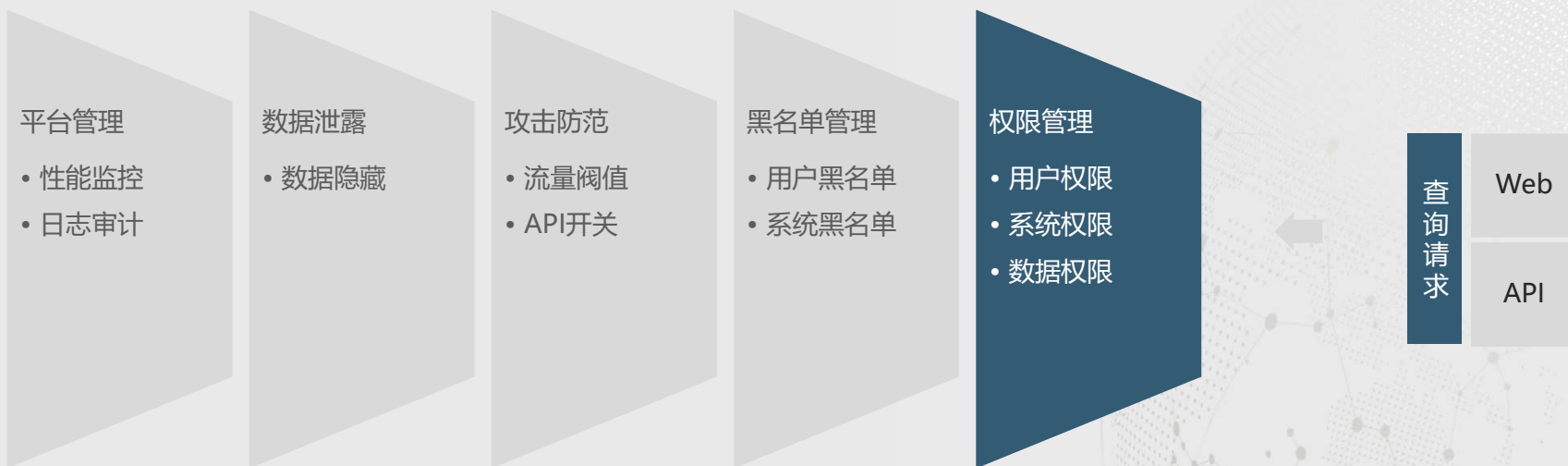
# 调度方案



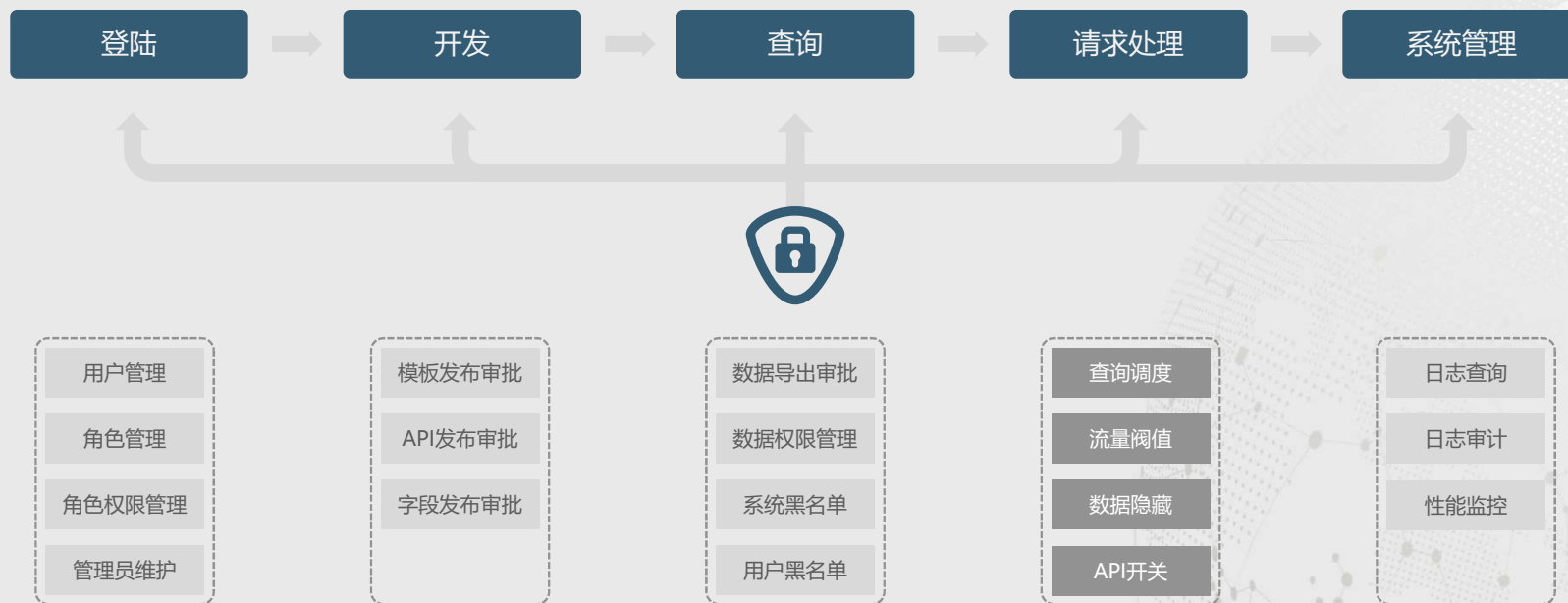
# 数据方案



# 安全机制



# 管理方案



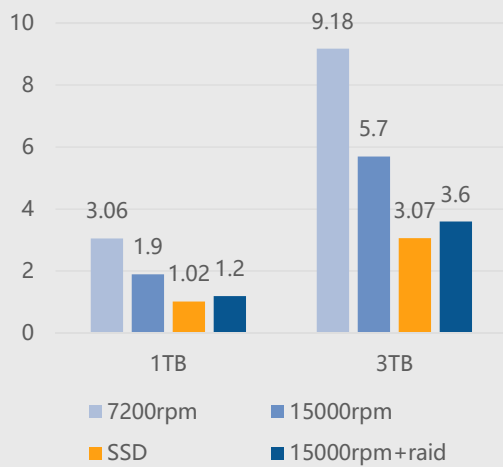
# 功能架构



# 性能评测

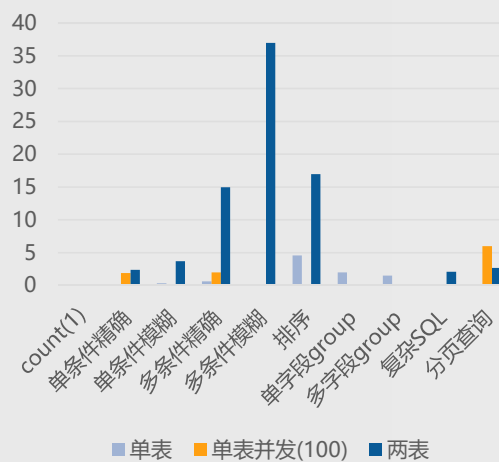
shop表, 数据量1TB, 50个字段  
 product表, 数据量3TB, 60个字段  
 6+1集群部署方案

建立索引时间 (小时)



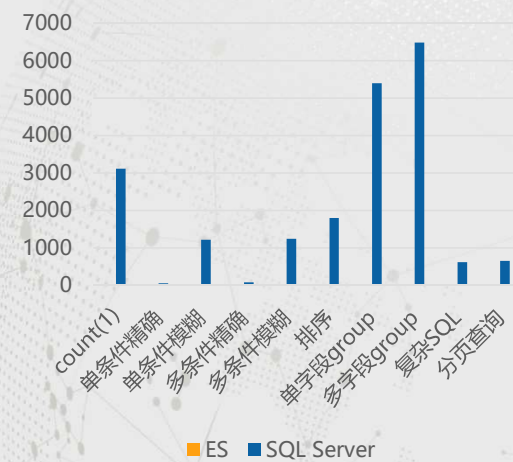
i5-6400 CPU @ 2.70GHz

查询性能 (秒)



16G 2133MHz DDR4

SQL Server 查询性能对比 (秒)



1T HDD 7200rpm

1000M

# 实施方案

1. 实施方案 2. 实施步骤 3. POC方案

# 实施方案



# 实施步骤

## 需求分析

- 时间：2w
- 业务需求分析
- 对接数据分析
- 对接系统分析
- 需求评审

## 方案设计

- 时间：2w
- 系统方案设计
- 测试方案设计
- 实施方案设计
- 设计评审

## 系统实施

- 时间：10w
- 系统开发
- 外部对接
- 数据源配置

## 测试验收

- 时间：6w
- 单元测试
- 联调测试
- 性能测试
- 验收测试

# POC方案

- 在poc时间段内对建立元数据对应关系，完成索引的建立
- 同时将已完成的查询demo连接上行方的poc数据，方便行方进行更简便的验证

8

9月上旬



时长3天



研发经理x1  
开发人员x1  
技术总监x1

仿真环境搭建

脱敏数据准备

系统部署

建立索引

查询验证



4核4线程 3.00GHz



16G 2133MHz DDR4



SSD 120G (系统盘)  
1T 7200rpm (数据盘)



1000M

专业 诚信 实干

Beyebe 比一比网络科技



感谢聆听

THANKS