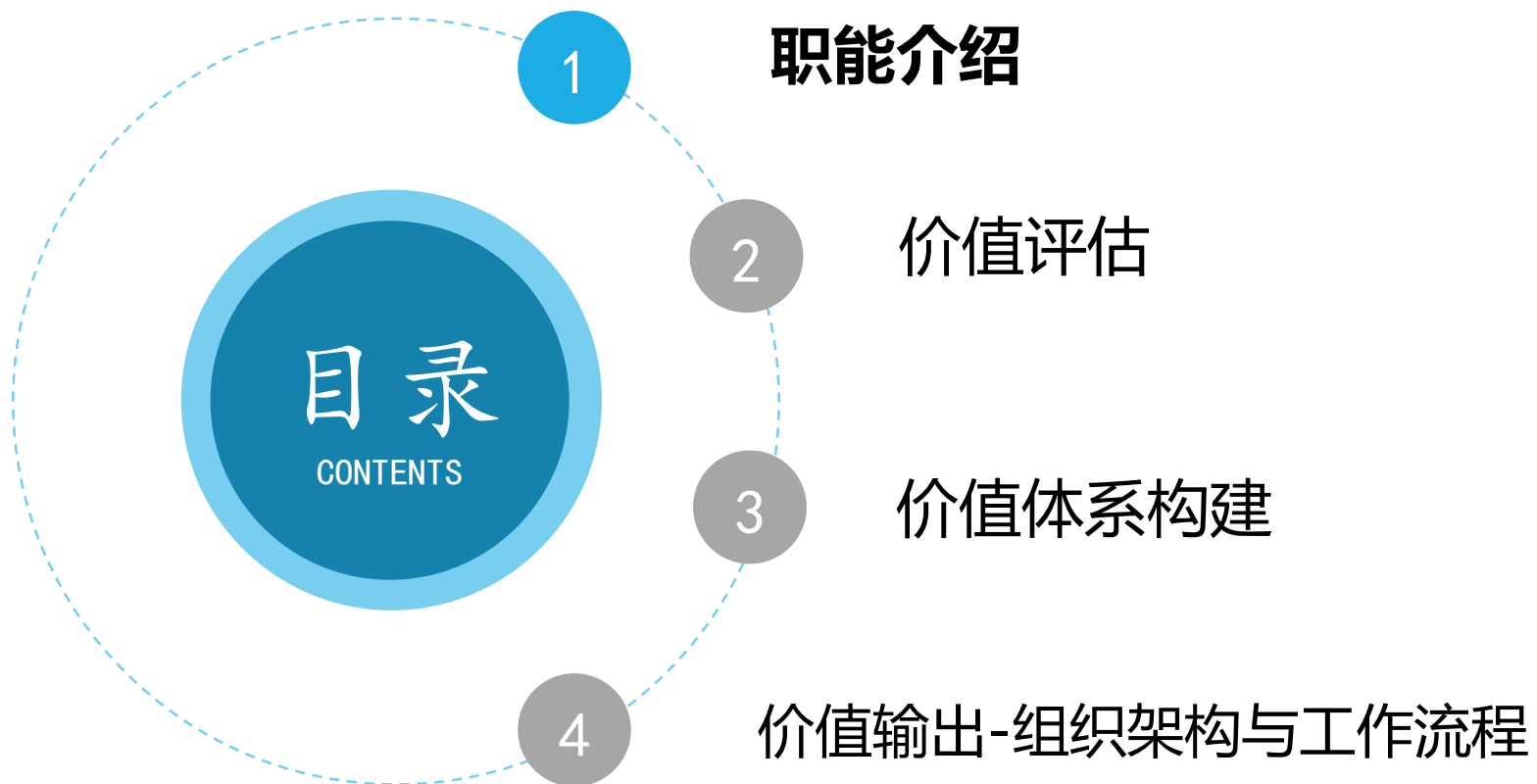




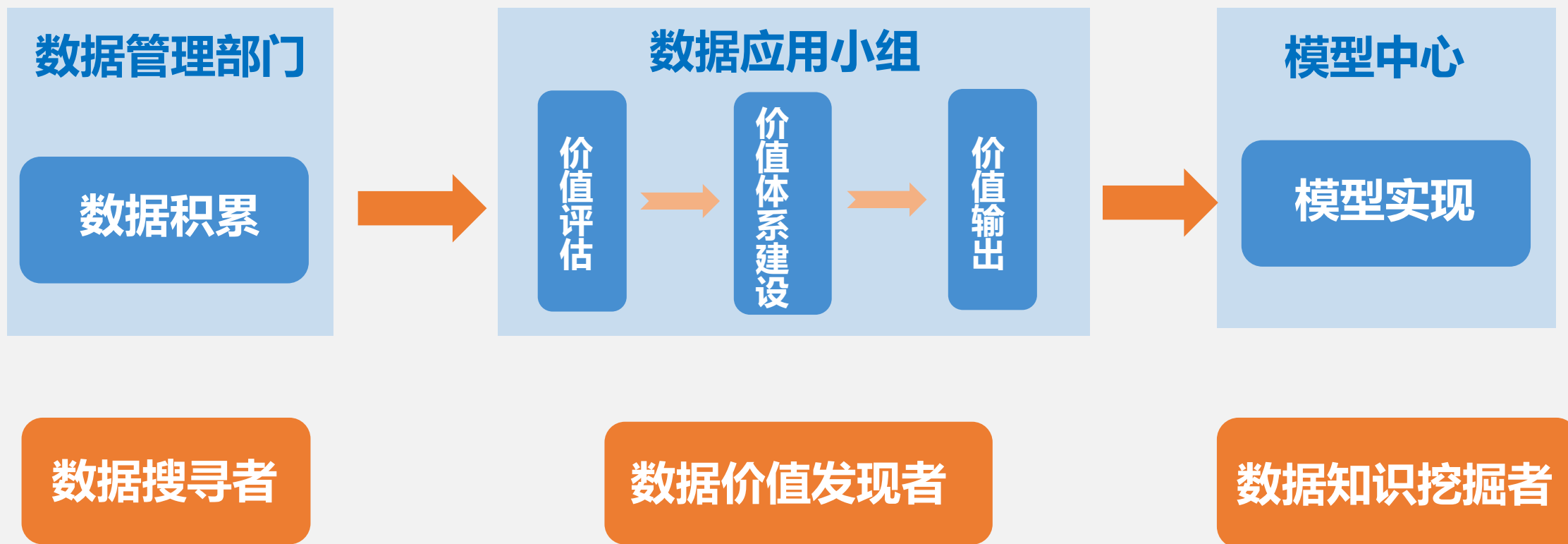
数据应用小组职能与工作流程介绍



大数据金融风险管理专家
BUSINESS BIG DATA



目标：统筹公司全集团数据需求，负责数据质量的把控和交付，逐步实现**数据与模型的一体化**。对公司已有和未来接入的数据进行数据来源检查、数据质量评估、对比数据相似度，评估数据是否满足业务需求、提出指标设计的建议，同时发掘数据指标可应用的场景、尝试数据可支持与应用的模型。





数据价值的发现者

服务介绍

服务内容



服务对象



价值评估



数据质量评估，数据价值评估，
可用场景评估



价值体系建设



全公司指标池建设，指标规范化推进，
指标的日常管理与维护



价值输出

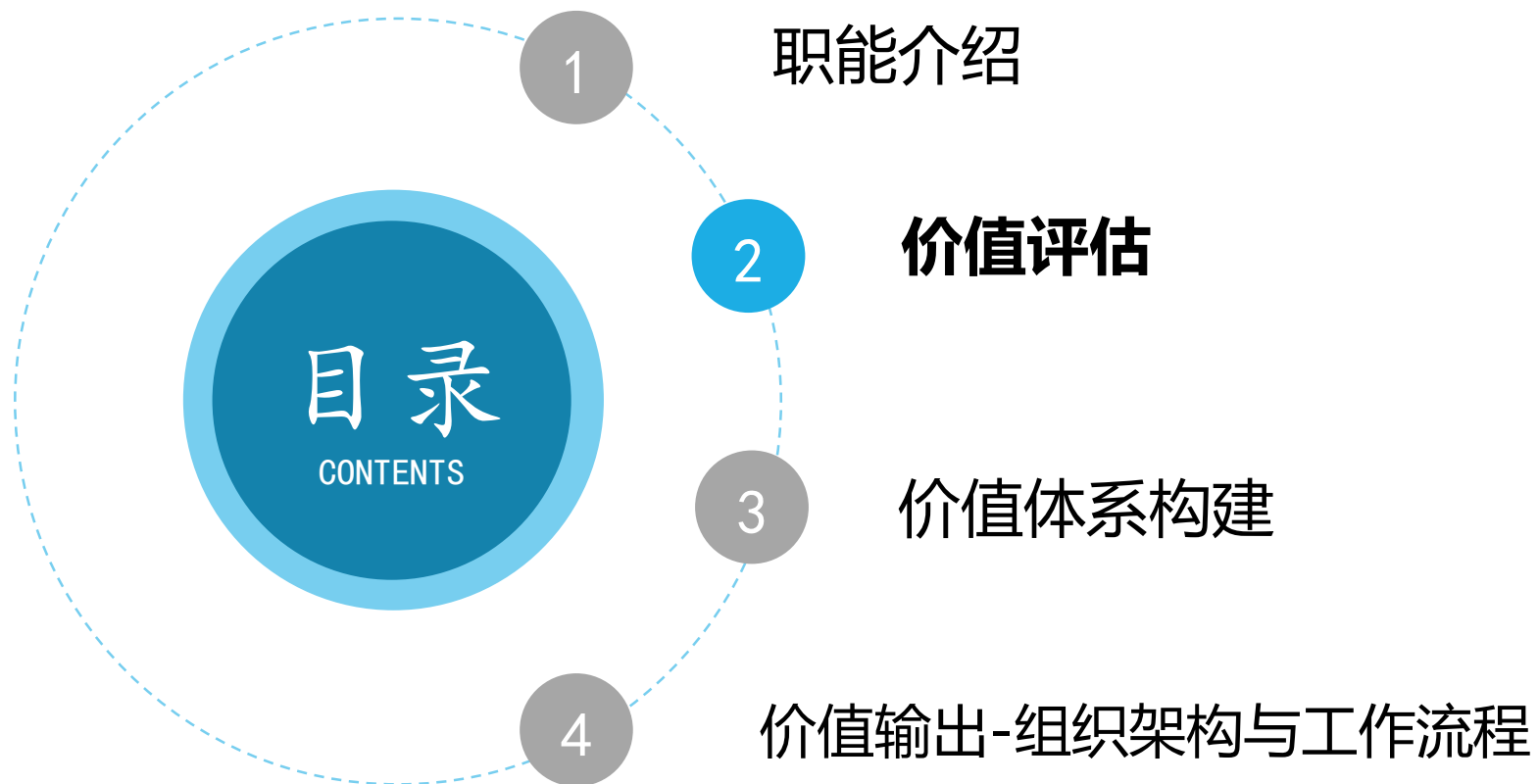


指标需求评估，指标结果输出

➢数据管理部门
➢产品人员

➢项目人员
➢产品人员
➢模型人员

➢项目人员
➢产品人员
➢任何有指标需求的同事







完整性

- **缺失率**
 - ① 字段的完整性
 - ② 数据的完整性（空值）
- **异常值**
 - ① 超出正常取值范围
 - ② 类型不相同



一致性

- **表间一致性**
相同数据来源的数据间对比
- **外部一致性**
与公司数据内容的对比



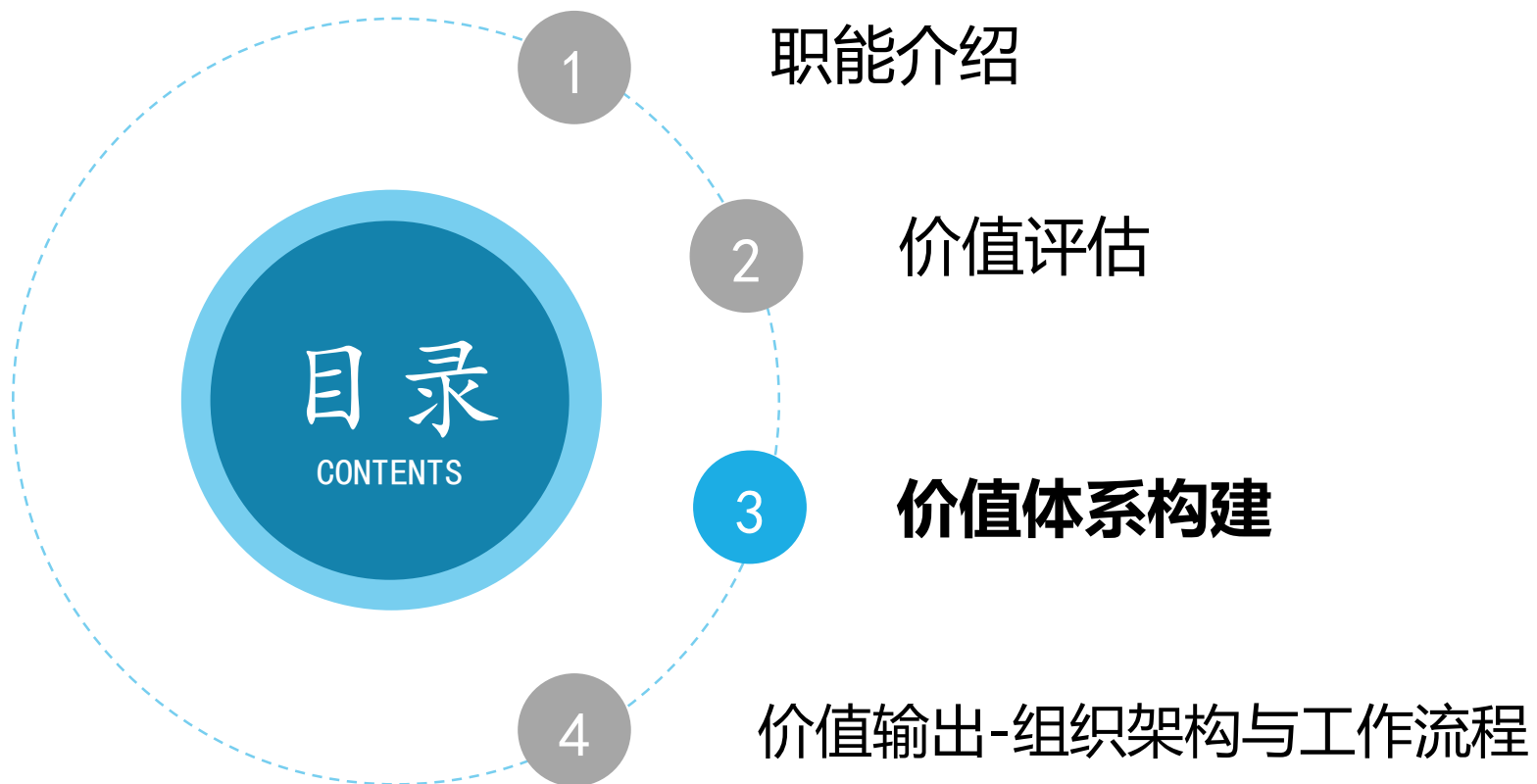
有效性

1. 重复率
2. 零值率
3. 数据类型
4. 内容规范



及时性

1. 全库更新频率
2. 表单更新频率
3. 字段更新频率





指标池是数据价值的载体，承载着对指标的所有需求，是数据价值直接的展示平台



指标池

痛点

- 相同指标重复开发
- 指标计算消耗资源且时间久
- 计算规则没有整合

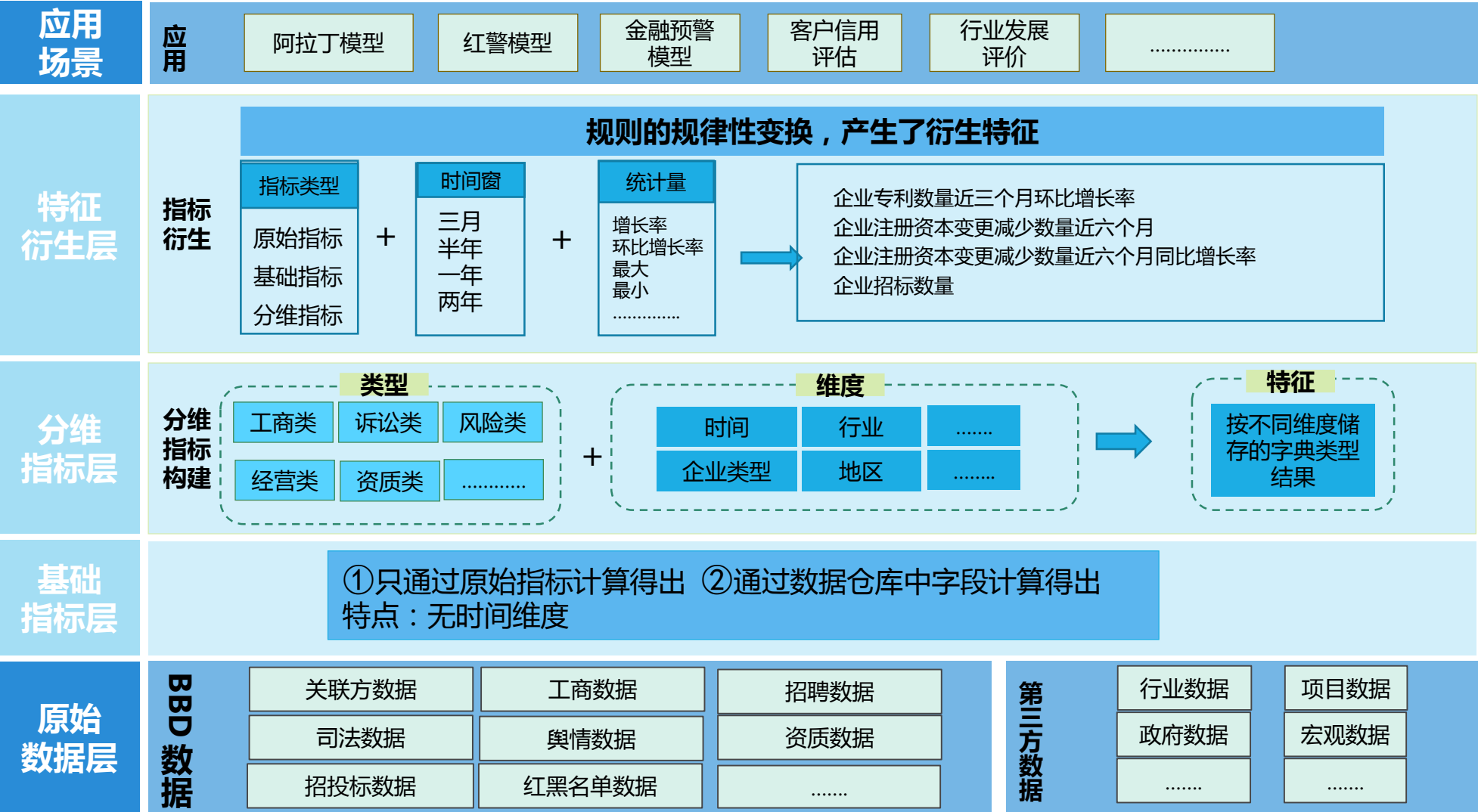
建设方向

- 对指标进行分表分类，全量指标结果的更新储存
- 指标构建规则的整合，达到指标复用
- 血缘管理，从数据定位到项目

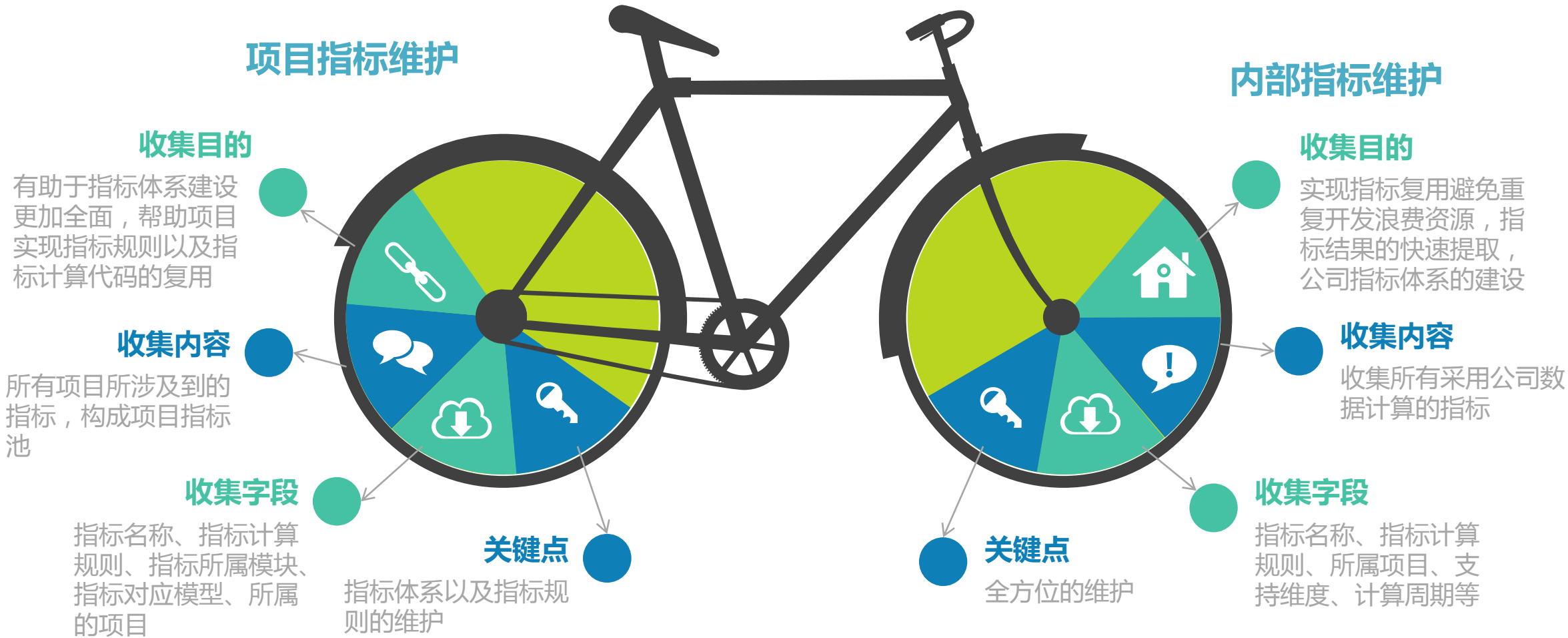
未来服务

- 对指标结果进行初步统计分析
- 提交名单，快速提取指标结果
- 指标计算逻辑的更新与维护
- 指标体系构建，对指标进行价值评估分类

指标池：用于存放所有指标其目的是统一出入口及标准，便于支持业务及上层数据查询，减少重复开发，提高效率。



基于所有项目的指标池扩充、丰富以及维护





① 平台建设目的

实现从数据--计算--指标--模型--项目的全流程追踪，做到指标的血缘管理

② 管理人员职责

- 定时更新指标
- 定时导入指标血缘关系
- 对出现问题的计算逻辑进行维护修改

③ 使用者权限

- 指标计算所用数据的查询
- 指标所属项目
- 所有指标名称

Apache Atlas

SEARCH CLASSIFICATION GLOSSARY

Basic ☒ Advanced ?

Search By Type

hive_table x

Search By Classification

Select Classification

Search By Term

Search Term

Search By Text

index_pools

Clear Search

Results for: (Type: hive_table) AND (Query: index_pools)

If you do not find the entity in search result below then you can [create new entity](#)

Showing 1 records From 1 - 25

☐ Exclude sub-type

<input type="checkbox"/>	Name	Owner	Description	Type
<input type="checkbox"/>	index_pools	phfz		hive_table

从字段realcap经过clu的计算得到指标qy_zzzb的结果

qy_zzzb (hive_column)

Classifications: original +

Term: +

Properties

Lineage

Relationships

Classifications

Audits

realcap

clu

qy_zzzb



一共有150个指标，2种计算方式

index_pools (hive_table)

Classifications: +

Term: +

Propagated Classifications: original

Properties

Lineage

Relationships

Classifications

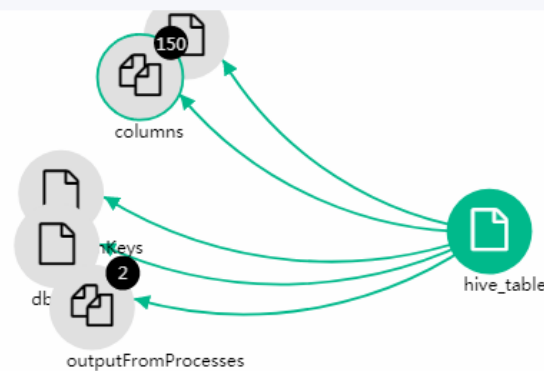
Audits

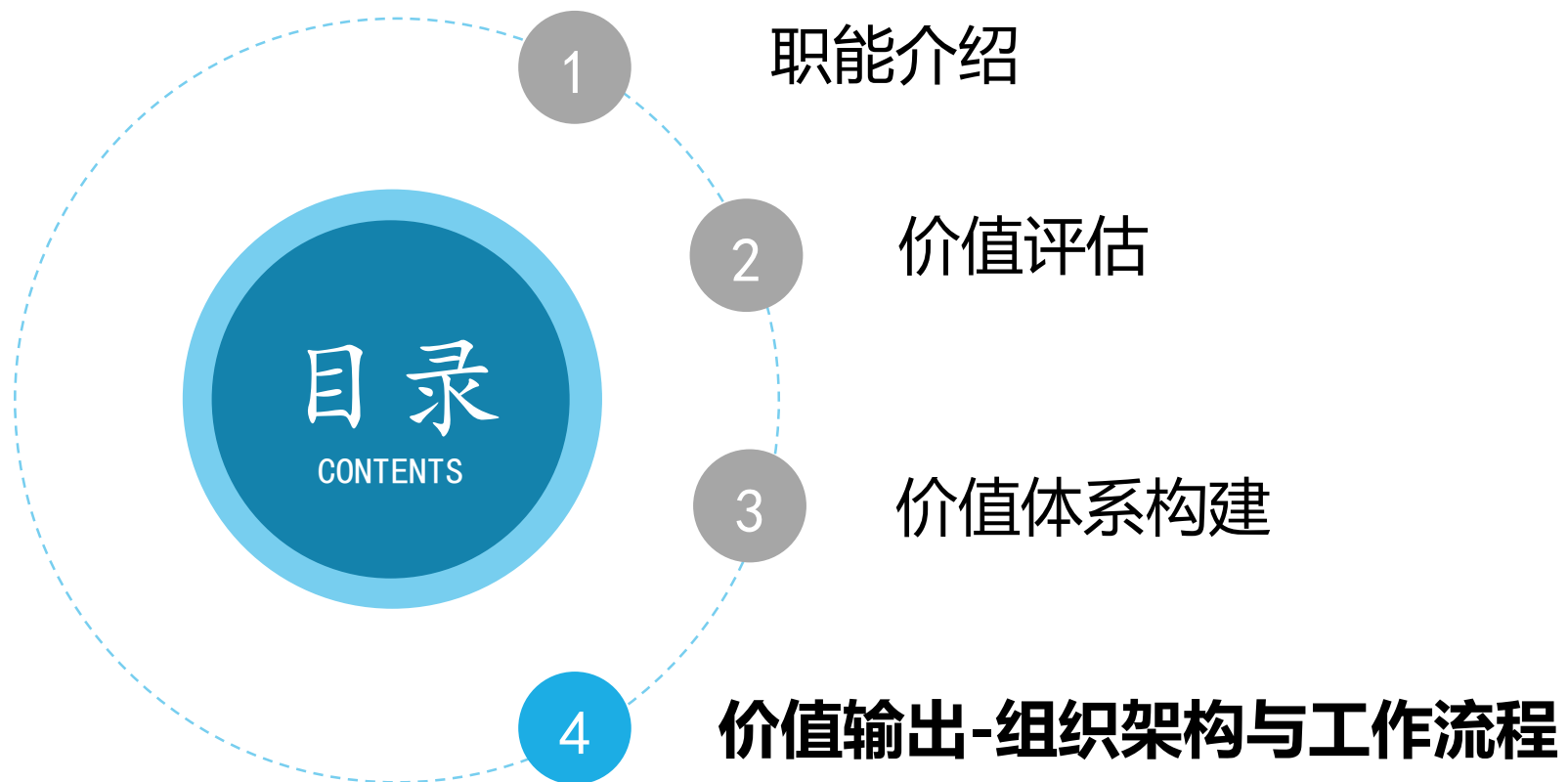
Schema

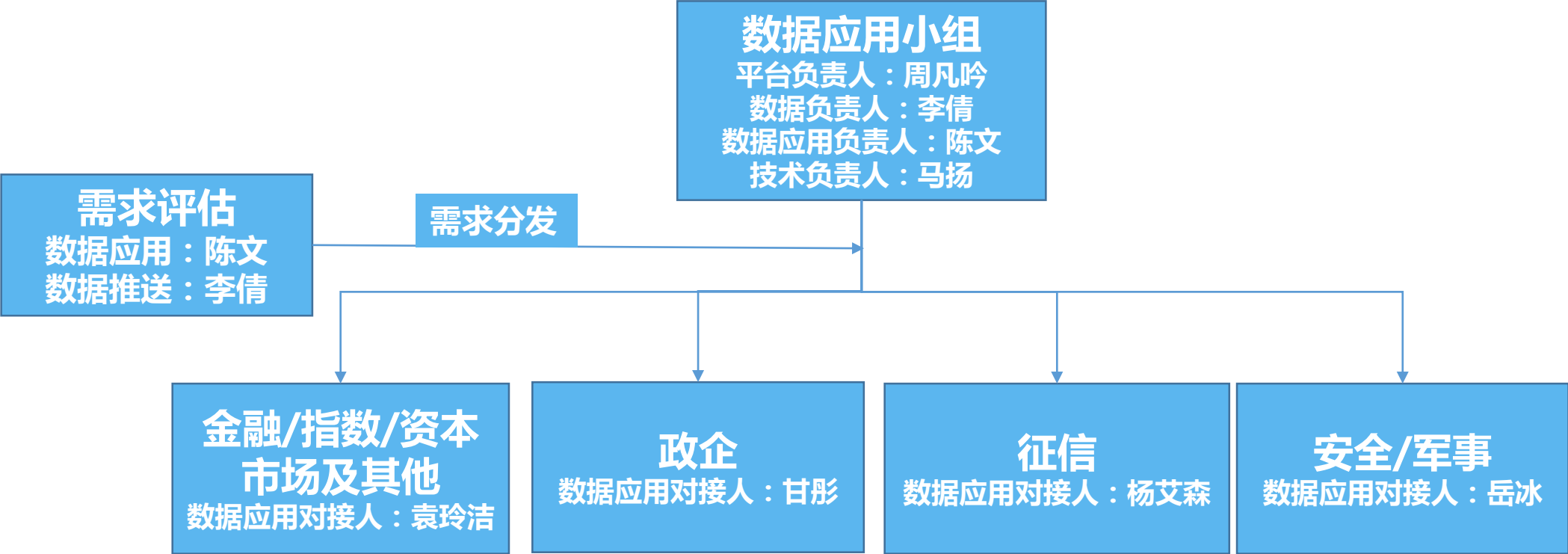
columns

Search Entities

1. bbd_qyxx_id (hive_column)
2. company_name (hive_column)
3. qy_bg_num_dim (hive_column)









平台负责人职责：

- 接纳外部需求和问题，进行初步分析（需求价值评估、紧急程度评估等）；
- 协调数据管理、数据应用、技术负责人进行进一步分析与需求评估；
- 协调发布需求、问题的进展和解决情况，以及平台需对外发布的其他信息。



数据应用负责人职责：

- 协助平台负责人进行数据应用、指标设计评估方面需求分析与评估；
- 对于数据应用/指标设计方面需求进行发布、问题的进展和解决情况跟进。



数据负责人职责：

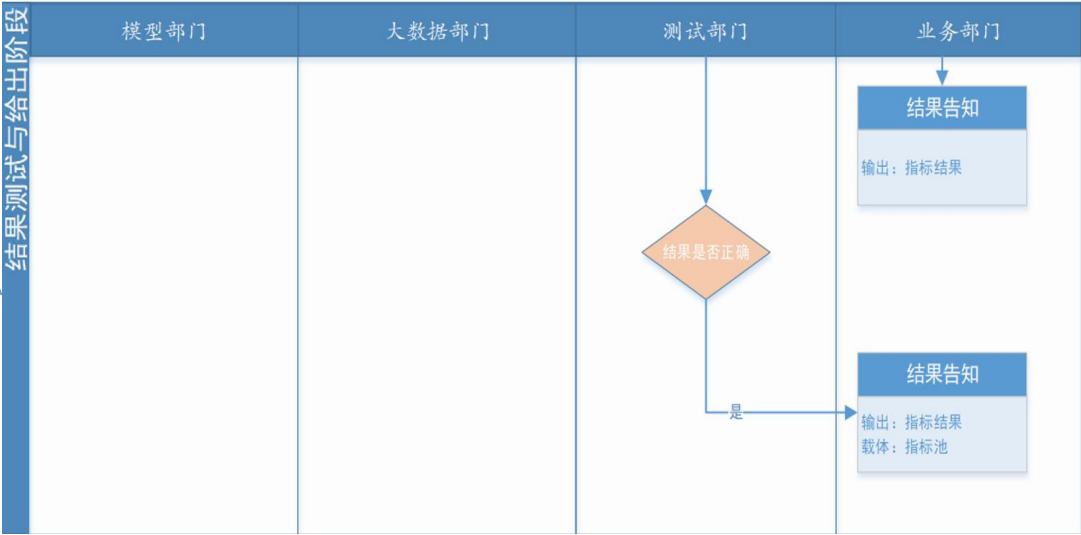
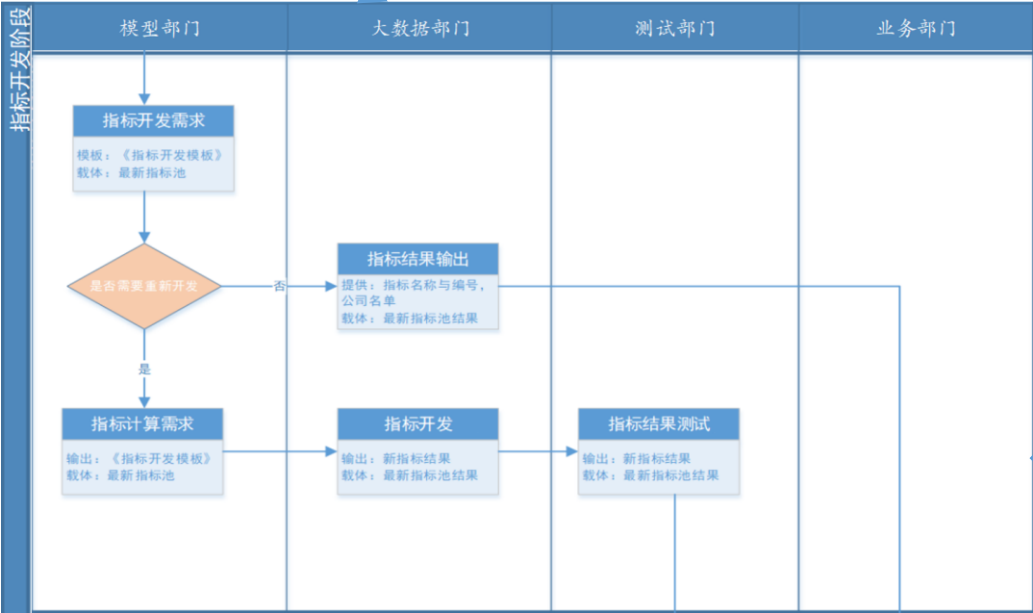
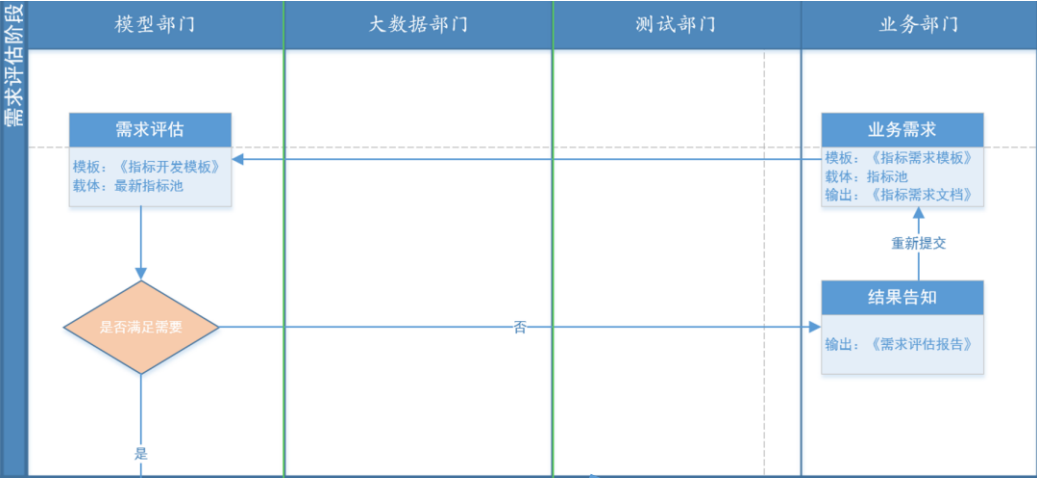
- 协助平台负责人进行数据管理方面需求分析与评估；
- 对于数据管理方面需求进行发布、问题的进展和解决情况跟进。

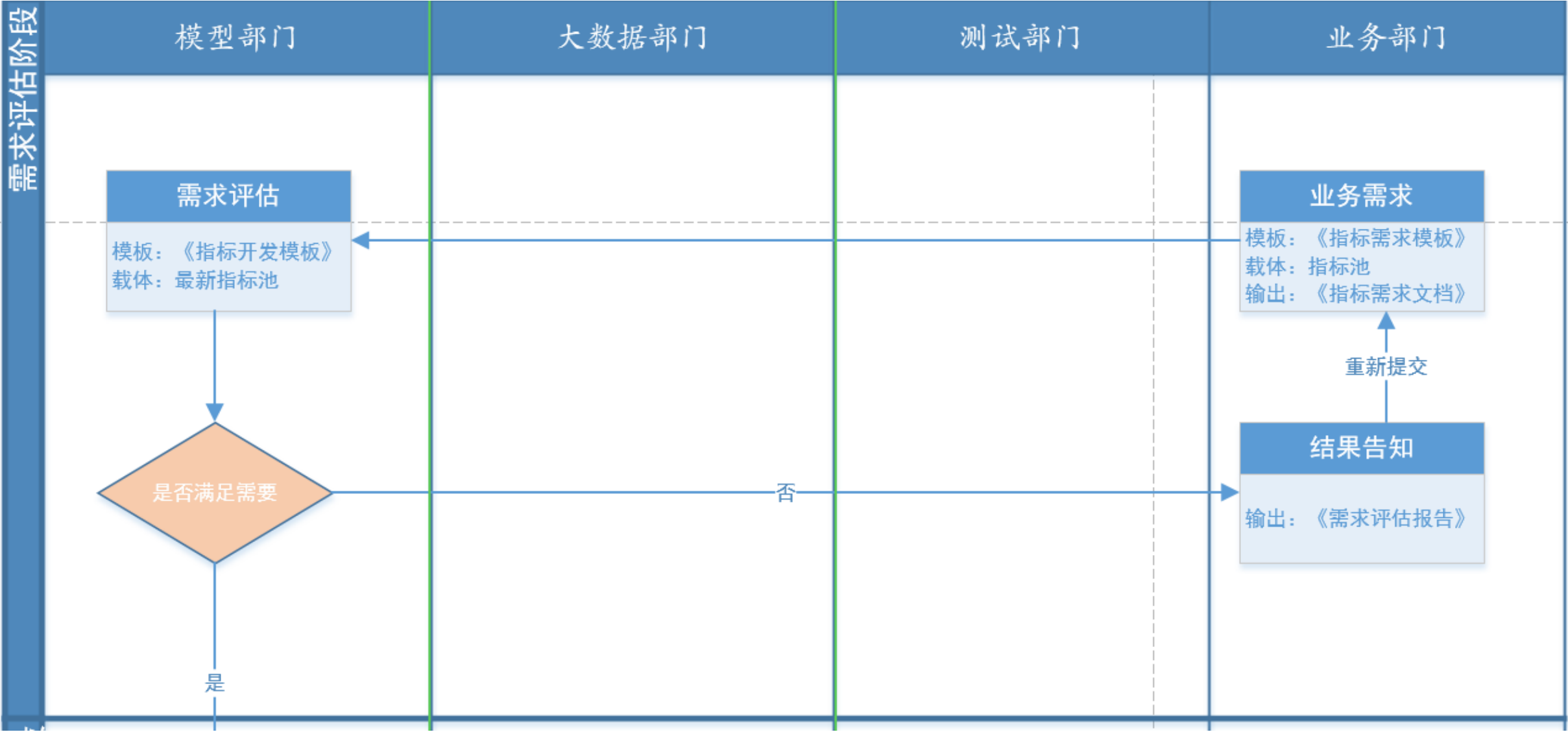


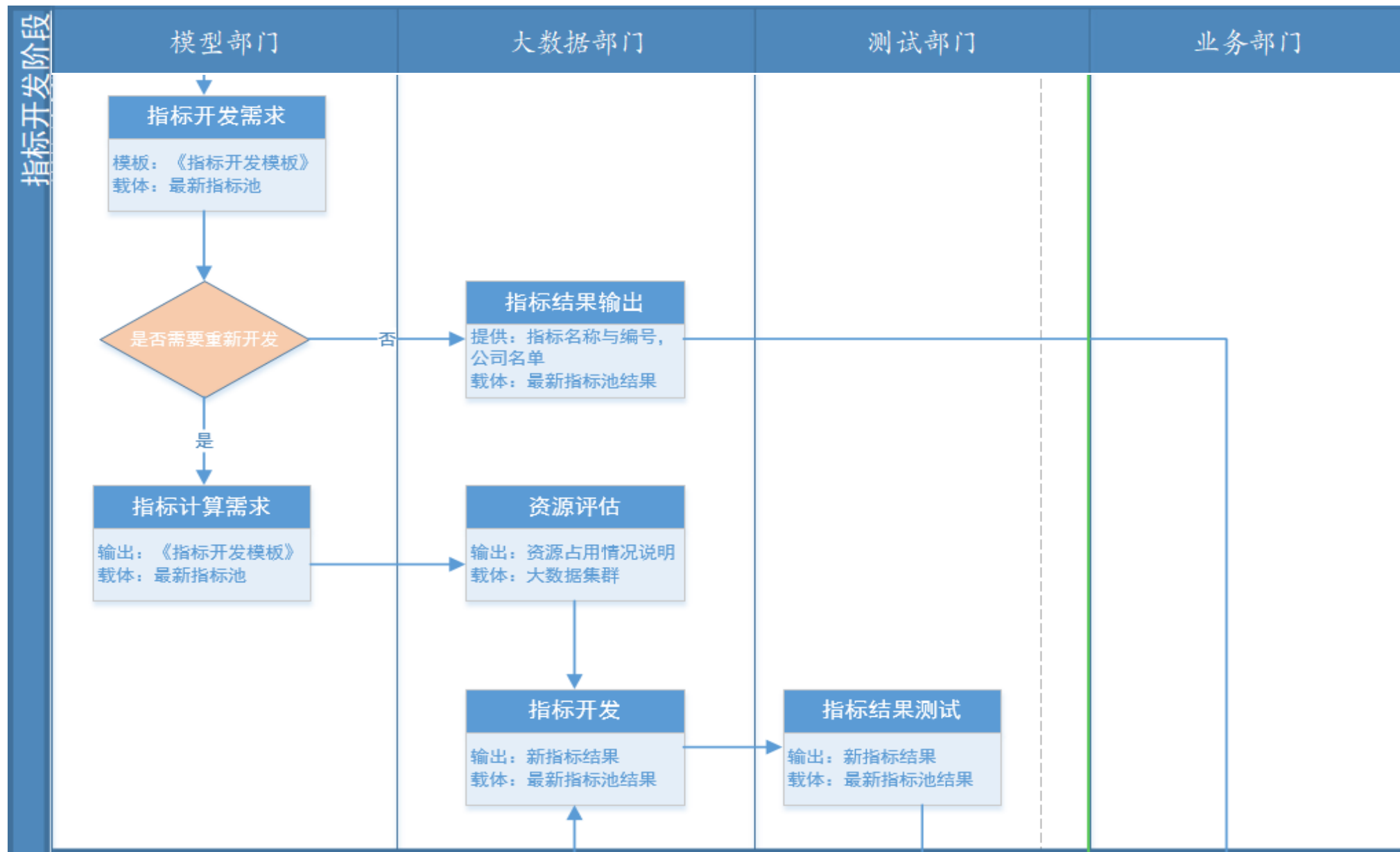
技术负责人职责：

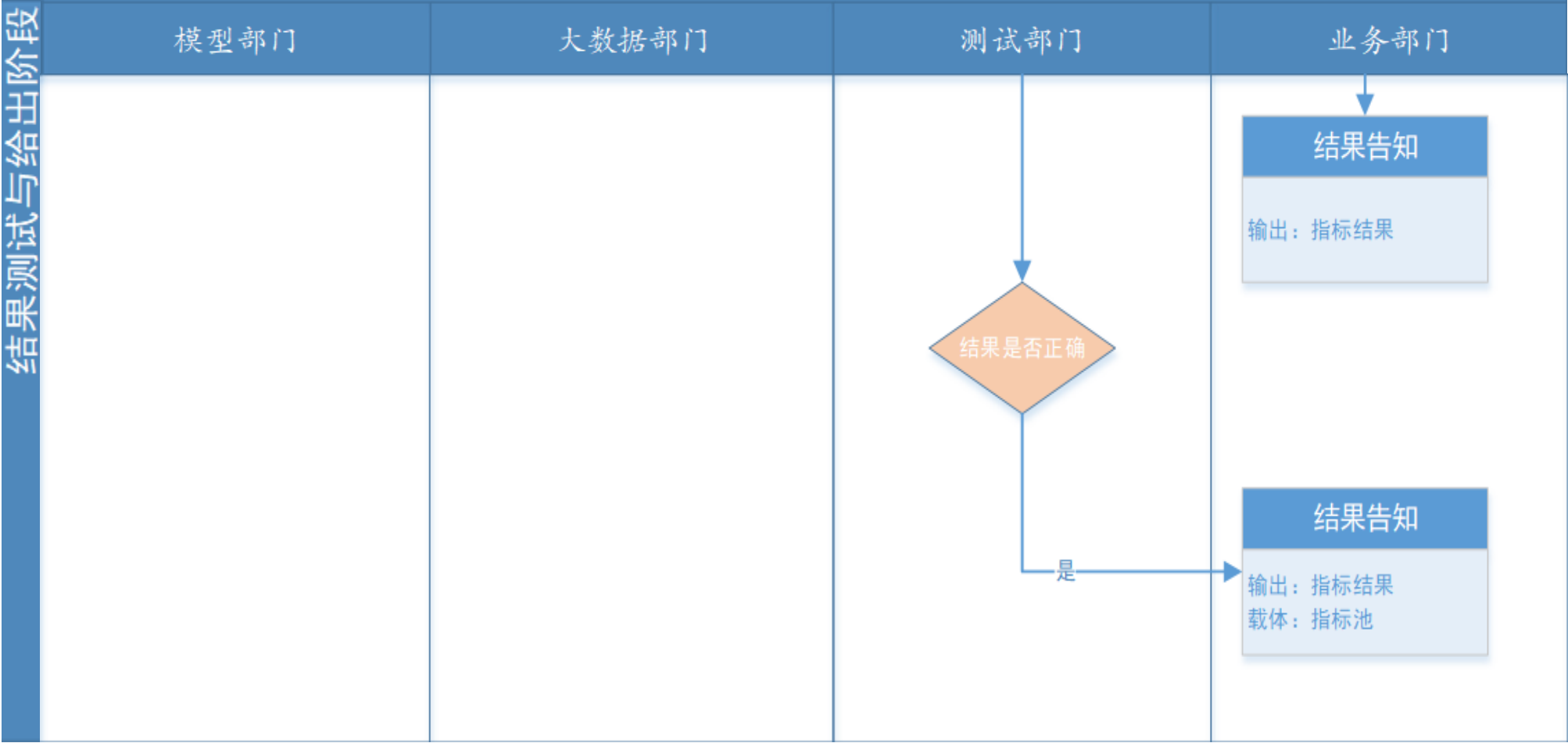
- 协助平台负责人进行大数据计算与指标开发方面需求分析与评估；
- 对于大数据计算与指标开发方面需求进行发布、问题的进展和解决情况跟进。

价值输出流程-数据应用评估流程









需求提交

提交文档
模板《指标需求模板》

- ◆ 模型：模型需求说明
- ◆ 模块：功能模块的设计与设计的功能点描述
- ◆ 指标：每个模块涉及的指标名称
- ◆ 数据：数据来源

需求评估

判断需求
评估指标是否满足模块的功能设计

- ◆ 存在性：需要指标是否存在，是否需要删除，更新指标库
- ◆ 必要性：新增指标是否可复用，开发的必要性

指标开发

计算开发
指标计算规则与代码

- ◆ 完成评估后，对需要新开发的指标进行规则编写，开发入库

指标结果的输出——数据价值的输出

A. 指标设计类

需求来源方主要为各业务部门产品经理和项目人员。

此类需求具体执行流程为：

- 1.禅道建立任务，同时发邮件给模型中心（**周凡吟、陈文**）、数据管理部（**李倩**）、大数据技术部（**马扬**）。
- 2.由李倩和陈文指定的负责人分发或拒绝需求。不确定的可以与李倩、陈文沟通确认。
- 3.若需求被采纳，后续时间安排和需求有两个部门合作安排与反馈。

B.指标结果提取类

需求来源方主要为模型中心及部门内部

此类需求具体执行流程为：

- 1.需求方在承接平台建任务。同时发邮件给模型中心（**周凡吟、陈文**）。
- 2.承接负责人发起评审，确认是否采纳需求。
- 3.若采纳，则纳入排期进入指标结果提取阶段。

C.指标开发与大规模计算类

需求来源方主要为模型中心及部门内部

此类需求具体执行流程为：

- 1.**发邮件**给模型中心（**周凡吟、陈文**）、数据管理部（**李倩**）、大数据技术部（**马扬**）。
- 2.承接任务的各部门负责人发起评审，确认是否采纳需求。
- 3.若采纳，则纳入排期进入指标开发阶段。

XXX项目

指标需求调研总表V1.0

是否立项
项目名称
项目编号
项管
产品经理

2019/xx/xx

A	B	C	D
变更日期	变更人	工作页名称	变更内容描述
2019/2/22	袁玲洁	3	模块2.4的指标新增了10个

A	B	C	D
功能模块名称	功能点	功能描述	模型需求说明
2-信用分析模型管理系统	2.1 信用监测预警模型	建设信用监测指标库、预警模型管理功能模块，通过对信用模型指标库和预警模型管理配置，实现对主体的信用监测预警模型以及可配置化	1、功能：需要可配置监测模型的指标，设置指标的阈值和预警条件，预警等级，更改模型后，及时生效，模型定期运行（每天）。需要通过海量数据运算和预警。包含了基础数据库和业务主题库的全库扫描和预警。 2、输出内容：输出预警的主体和预警的风险详细信息。包括自然人和企业主体。预警时间、预警类型、预警风险。 3、思路：目前监测预警指标明细和预警规则未确定。12月15日上线建议将评估模型的计算结果部分指标、以及明确的黑名单规则、企业行政处罚类的信息作为监测指标。
	2.2		
3-重点领域信用分析系统	3.4		
6-监测预警分析系统	6.3		
	6.5		
7-信用评价子系统	7.1		
8-信用舆情分析系统	8.5		
	8.6		

需求模板说明

功能模块编号	指标中文名称	类别	所用表与字段	计算逻辑	计算规则（选填）	数据来源	结果类型	备注
2.1	企业被告裁判文书数量	诉讼	[zgcpwsw](company_name,bbd_qyxx_id,casecode,sentence_date)	判断目标企业历史上所有作为被告涉及到的裁判文书数量	计算目标企业作为被告所涉及的裁判文书数量，对company_name计数，用casecode去重并且按照sentence_date判别裁判文书的时间	BBD数据仓库	整数型	
	企业被告开庭公告数量	诉讼	[ktgg](company_name_,case_code,trial_date)	判断目标企业历史上所有涉及到的裁判文书数量	计算目标企业作为被告所涉及的开庭公告数量，对compan_name计数，用case_code去重并且按照trial_date判别开庭公告的时间	BBD数据仓库	整数型	
2.2	企业被执行金额数量	诉讼	[zhixing](bbd_qyxx_id,exec_subject,case_create_time)	目标企业历史上所有被执行金额的和	计算目标企业涉及被执行的金额，对bbd_qyxx_id计数，用exec_subject求和并且按照case_create_time判别被执行公告的时间	BBD数据仓库	整数型	
	企业被执行数量	诉讼	[zhixing](bbd_qyxx_id,case_code,case_create_time)	目标企业历史上被执行的次数求和	计算目标企业涉及被执行总次数，对bbd_qyxx_id计数，用case_code去重并且按照case_create_time判别被执行公告的时间	BBD数据仓库	整数型	
3.4	企业变更数量	工商	[qyxx_bgxx](company_name,change_data)	目标企业历史上的所有变更次数求和	计算目标企业进行企业变更的数量，对bbd_qyxx_id计数，按照change_date判别变更的时间	BBD数据仓库	整数型	
	企业裁判文书数量	诉讼	[zgcpwsw](bbd_qyxx_id,casecode,sentence_date)	判断目标企业历史上所有作为被告涉及到的裁判文书数量	计算目标企业所涉及的裁判文书数量，对bbd_qyxx_id计数，用casecode去重并且按照sentence_date判别裁判文书的时间	BBD数据仓库	整数型	
	企业大学招聘数量	工商	[recruit](bbd_qyxx_id,bbd_recruit_num,pubdate,education_required)	判断目标企业招聘的大学本科的人数	计算目标企业大学招聘的数量，通过关键词“本科”查找大学招聘的目标公司,对bbd_qyxx_id计数，用bbd_recruit_num求和并且按照pubdate判别招聘的时间	BBD数据仓库	整数型	



成 都 · 贵 阳 · 北 京 · 合 肥 · 南 京 · 重 庆 · 新 加 坡 · 伦 敦

www.bbdservice.com

400-992-0866

感谢聆听, 数联铭品版权所有