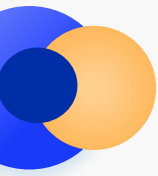


— 2023 —

# 数仓 Onedata 体系建设方法论

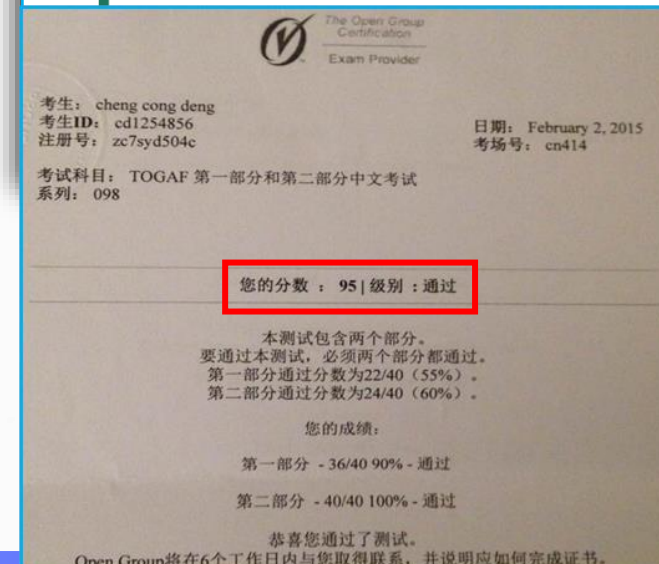
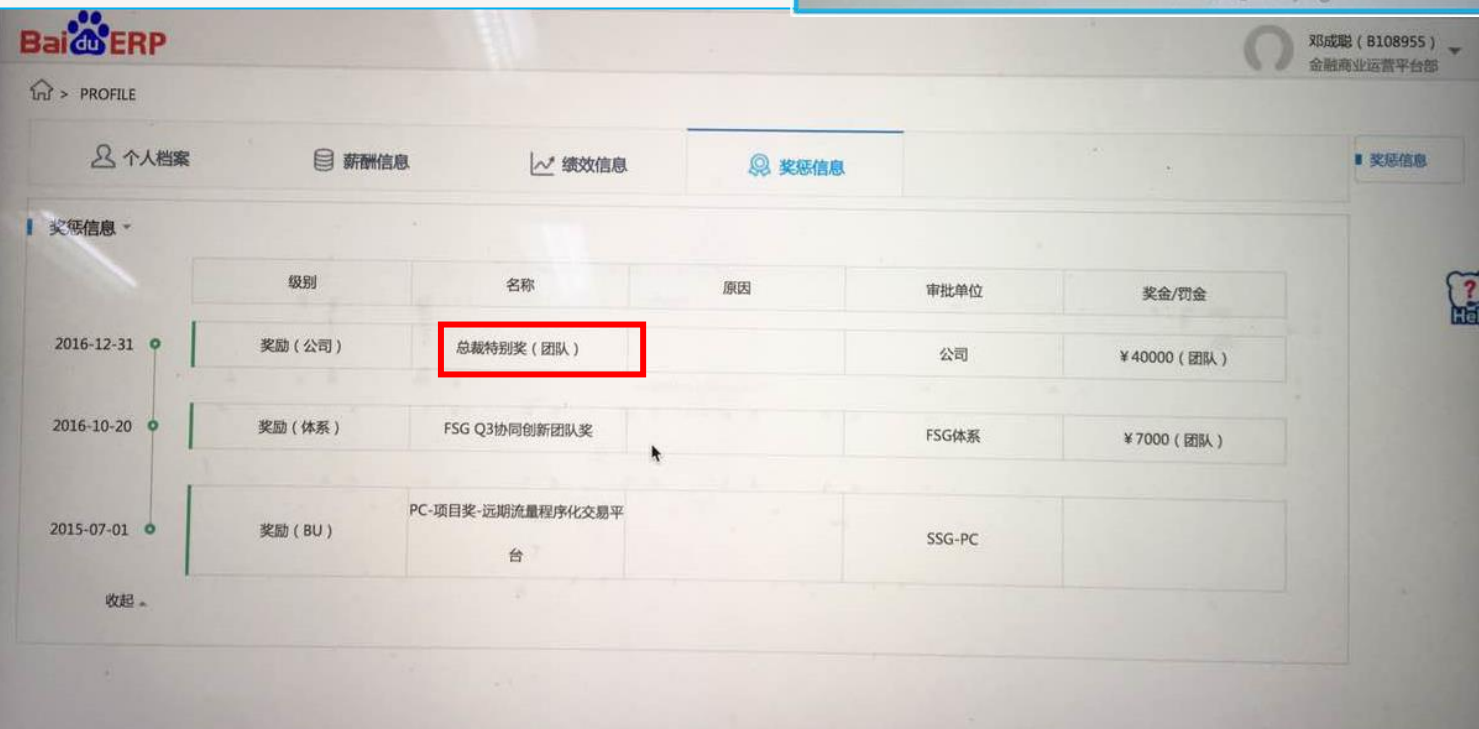


演讲人：邓成聪



# 讲师介绍

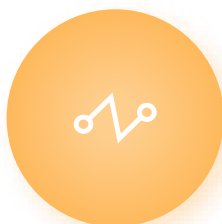
- 邓成聪，2000年毕业于武大大学，层服务于IBM，埃森哲，百度和华为，邓成聪是前百度大数据资深产品设计师、华为AI平台高端产品专家，一手孵化了华为ModelArts(获得斯坦福大学及全球智博会双料冠军)，曾代表华为参与编写中国人工智能平台标准。
- 邓成聪也是TOGAF注册企业架构师（95分，全球历史最高分）。
- 邓成聪在数据建模领域有20年经验，曾获百度总裁特别奖、华为总裁奖。



# CONTENTS 目录



方法论体系



数据建模  
流程工艺



实践案例



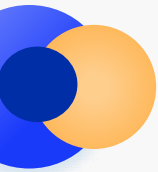
心得总结

# 01

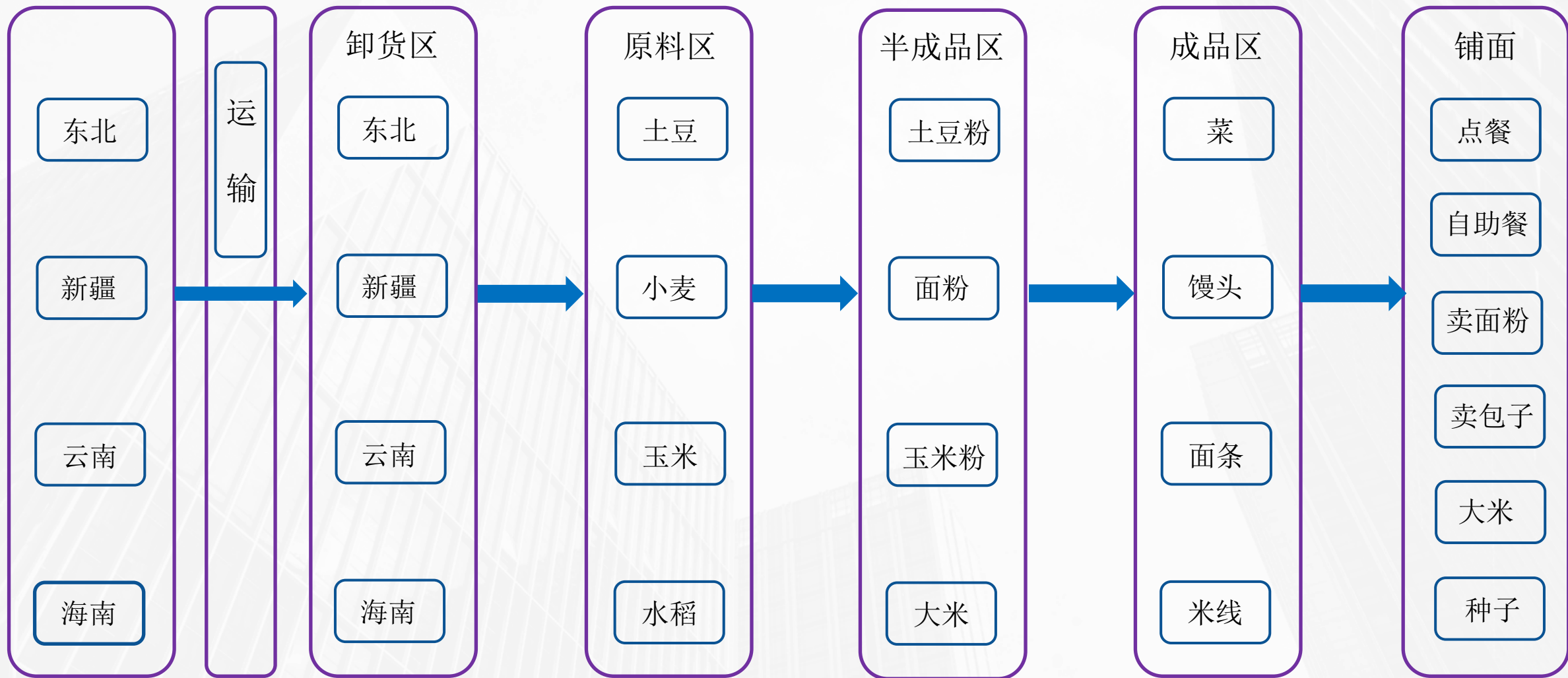
## 方法论体系

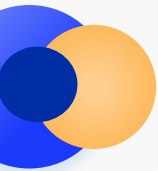




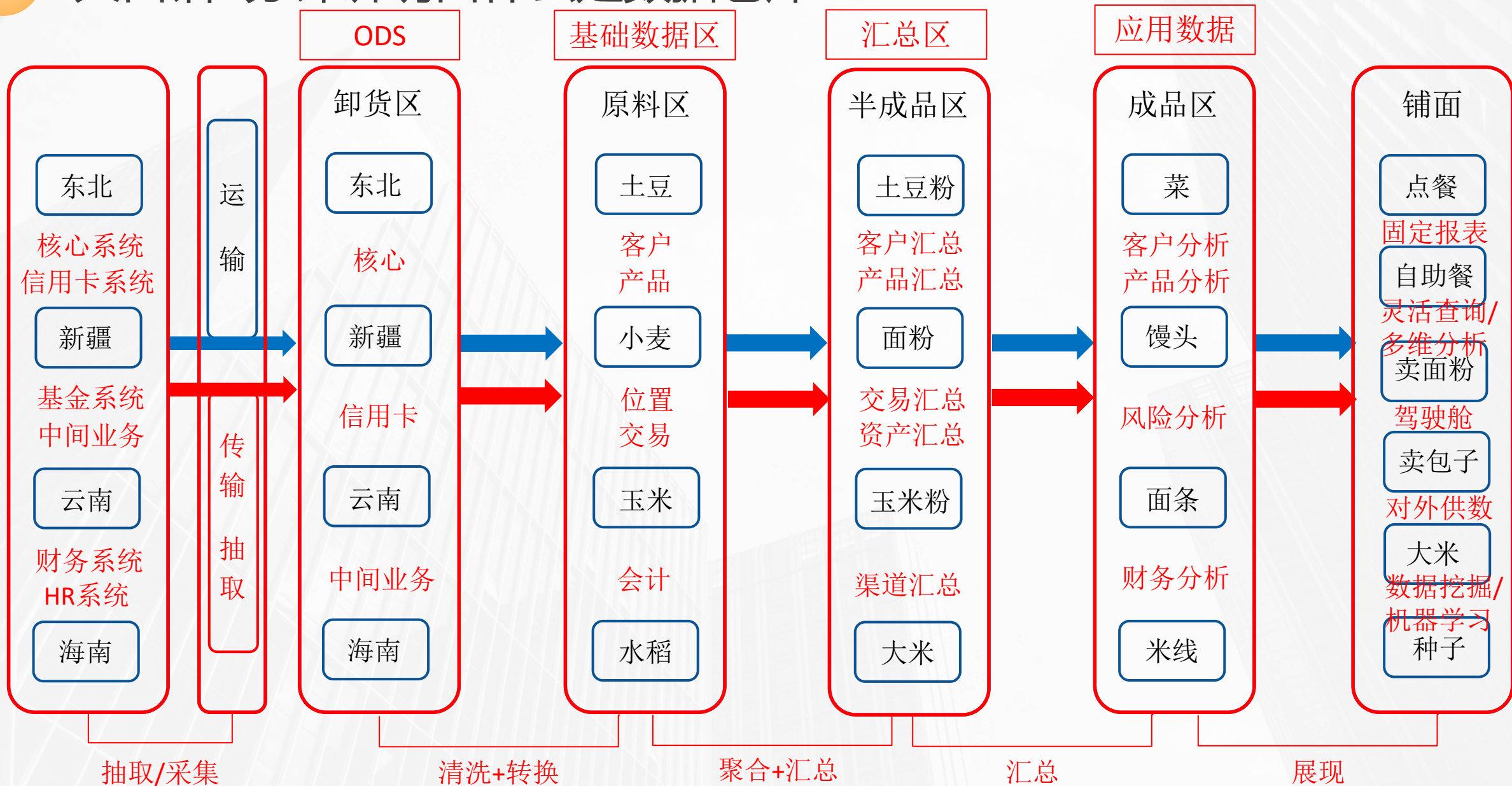


# 大白话3分钟讲明白什么是数据仓库





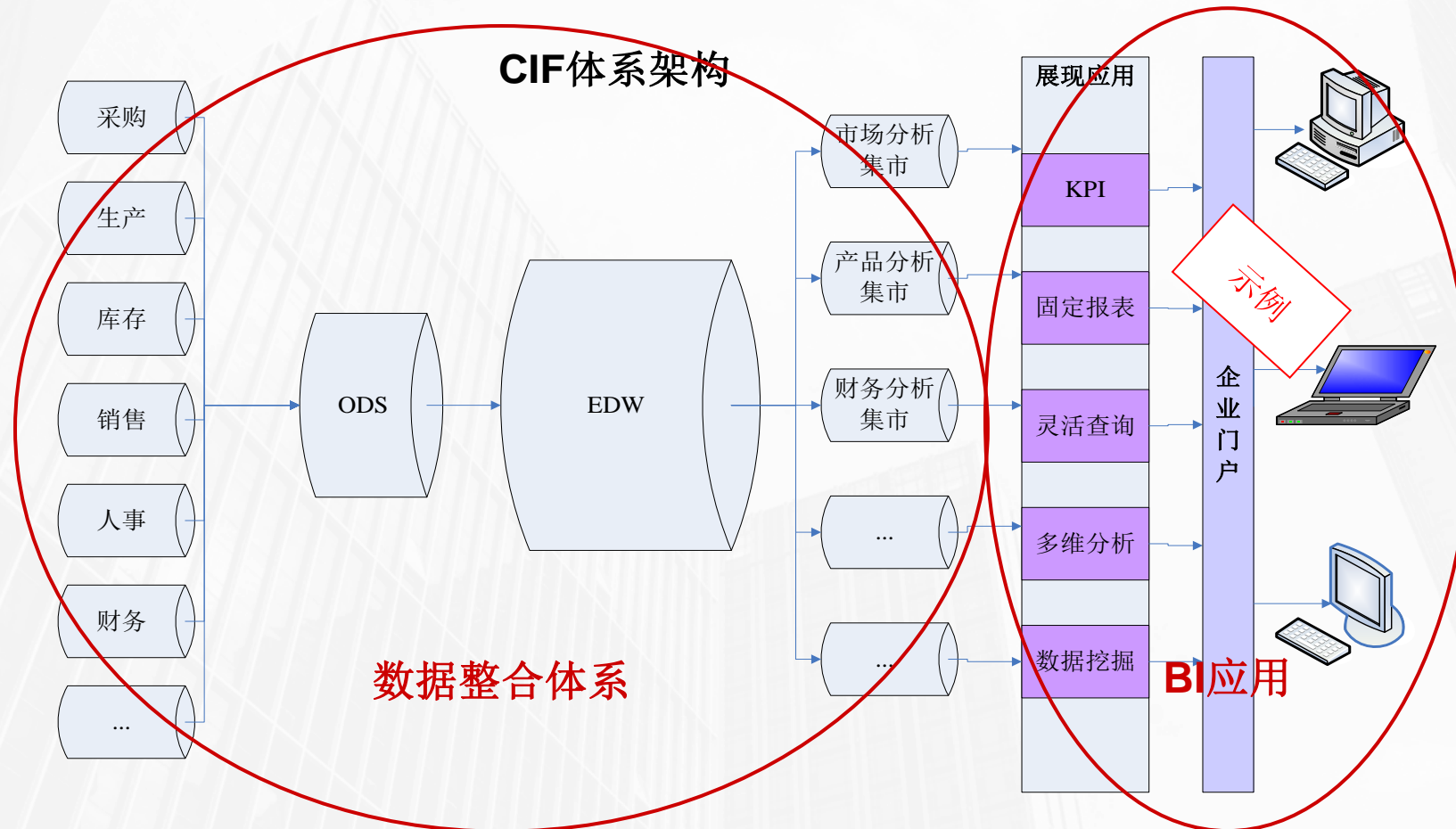
# 大白话3分钟讲明白什么是数据仓库



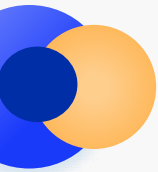


# 数据仓库方法论

- 在数据仓库领域的方法论，有两派理论，分别由Bill Inmon及Ralph Kimball所倡导；
- 在Bill Inmon新的方法论体系中，融合了两派方法论，提出了CIF（Corporate Information Factory）的体系架构，很多企业的数据仓库及BI架构，都遵循了这种架构理论。
- 目前在大型企业，成熟的数据仓库解决方案中，也最终走向了CIF的体系架构。

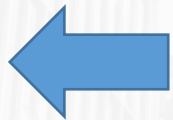


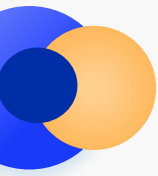




# 数据模型的地位

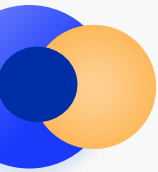
- 在数据仓库中，数据模型的地位就像发动机一样重要。
- 数据模型的稳定性、可获展性、数据模型易用性等，是评价建模水平的关键标准。建模中需要有相应的策略去指导。





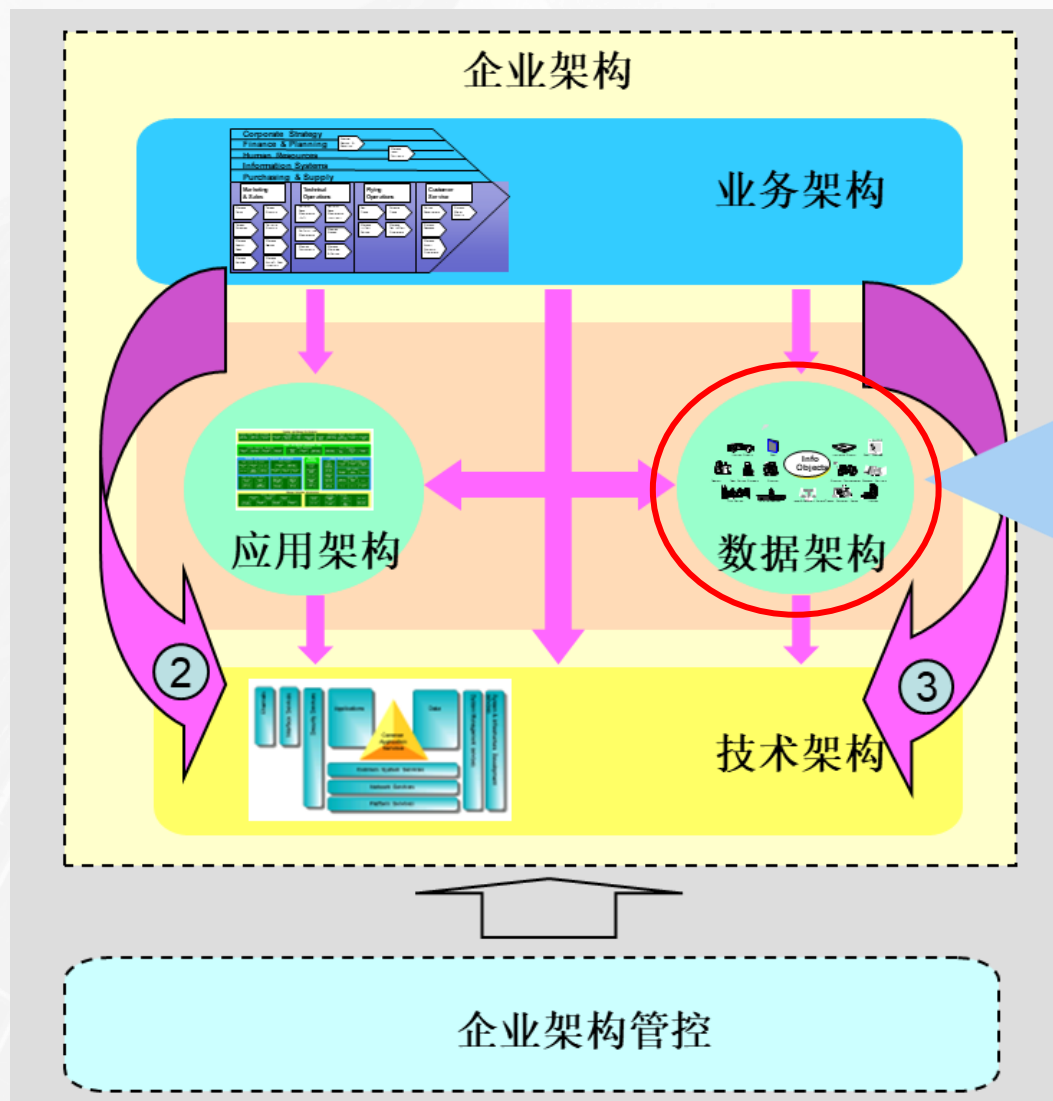
# Onedata体系

- Onedata 概念最初是由阿里巴巴提出
- OneData的核心，本质是数据在数据仓库里面，全局完整、含义一致、避免重复建设。Onedata的理念与Bill inmon的理念，企业架构的思想理念其实是一致的，可以说殊途同归！
- Onedata 建设的从业务架构设计到模型设计，从数据研发到数据服务，做到数据可管理、可追溯、可规避重复建设。即数据只建设一次。
- OneData又主要抽象成三个部分，分别是：OneID、OneModel、OneService。
  - ✓ 第一部分：OneModel 致力于实现数据的标准与统一；
  - ✓ 第二部分：OneID 致力于实现实体的统一，让数据融通而非以孤岛存在，为精准的用户画像提供基础；
  - ✓ 第三部分：OneService 致力于实现数据服务统一，让数据复用而非复制。

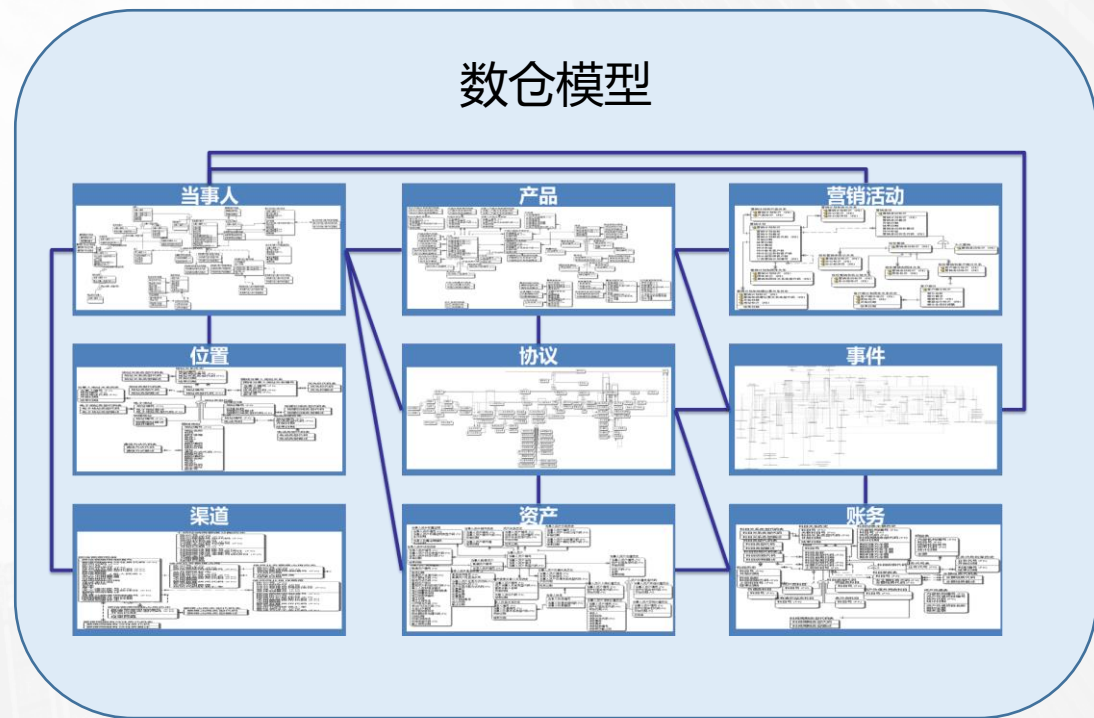


# 企业架构与数据仓库的关系

- 在企业架构的数据架构中，企业级数据模型会落地在数据仓库中，也是数仓的参考模型



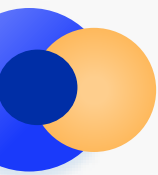
展开



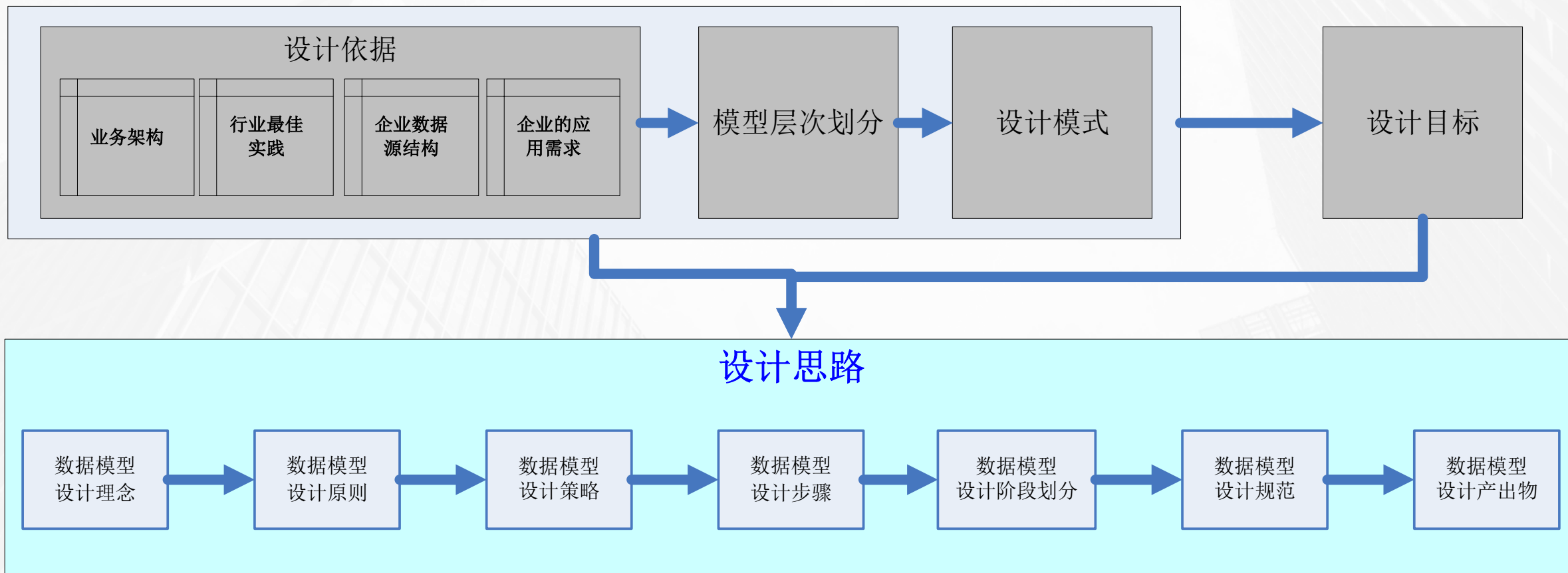
# 02

## 模型设计工 艺流程

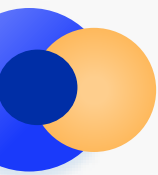




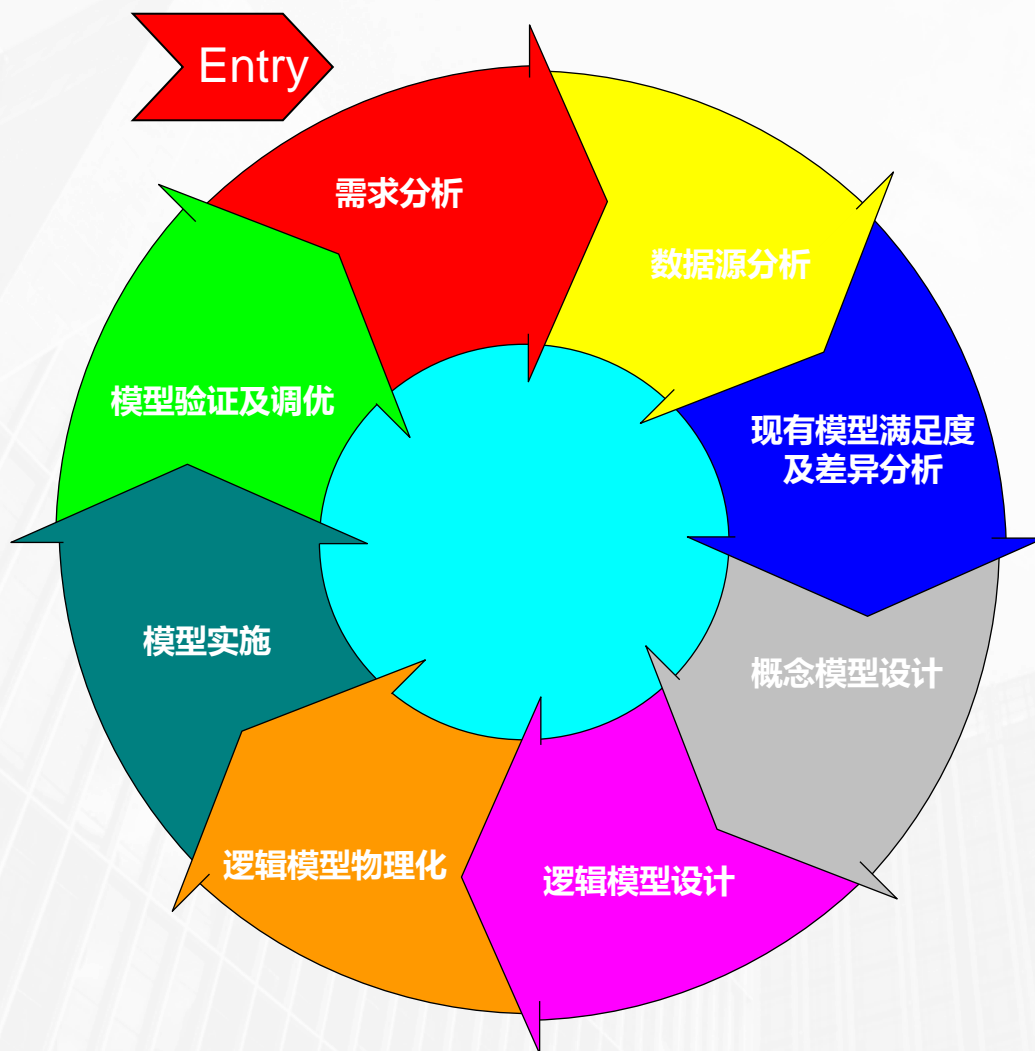
# Onedata数据模型设计思路







# 数据仓库数据模型设计流程



- 8个过程
- 过程可定义
- 迭代开发
- 参与角色
  - 业务分析师 (BA)
  - 模型设计师
  - DBA
  - ETL工程师

03

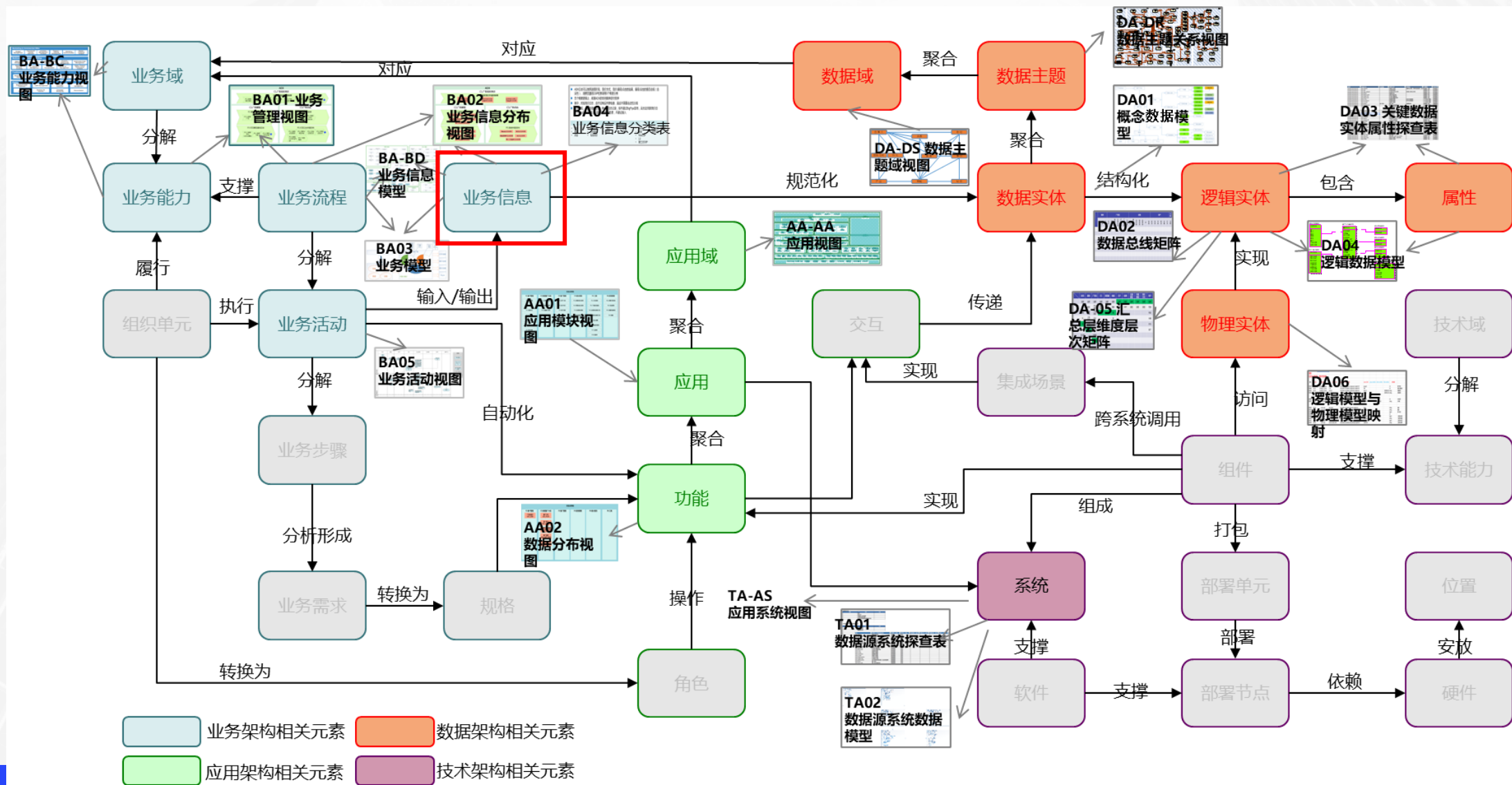
最佳实践





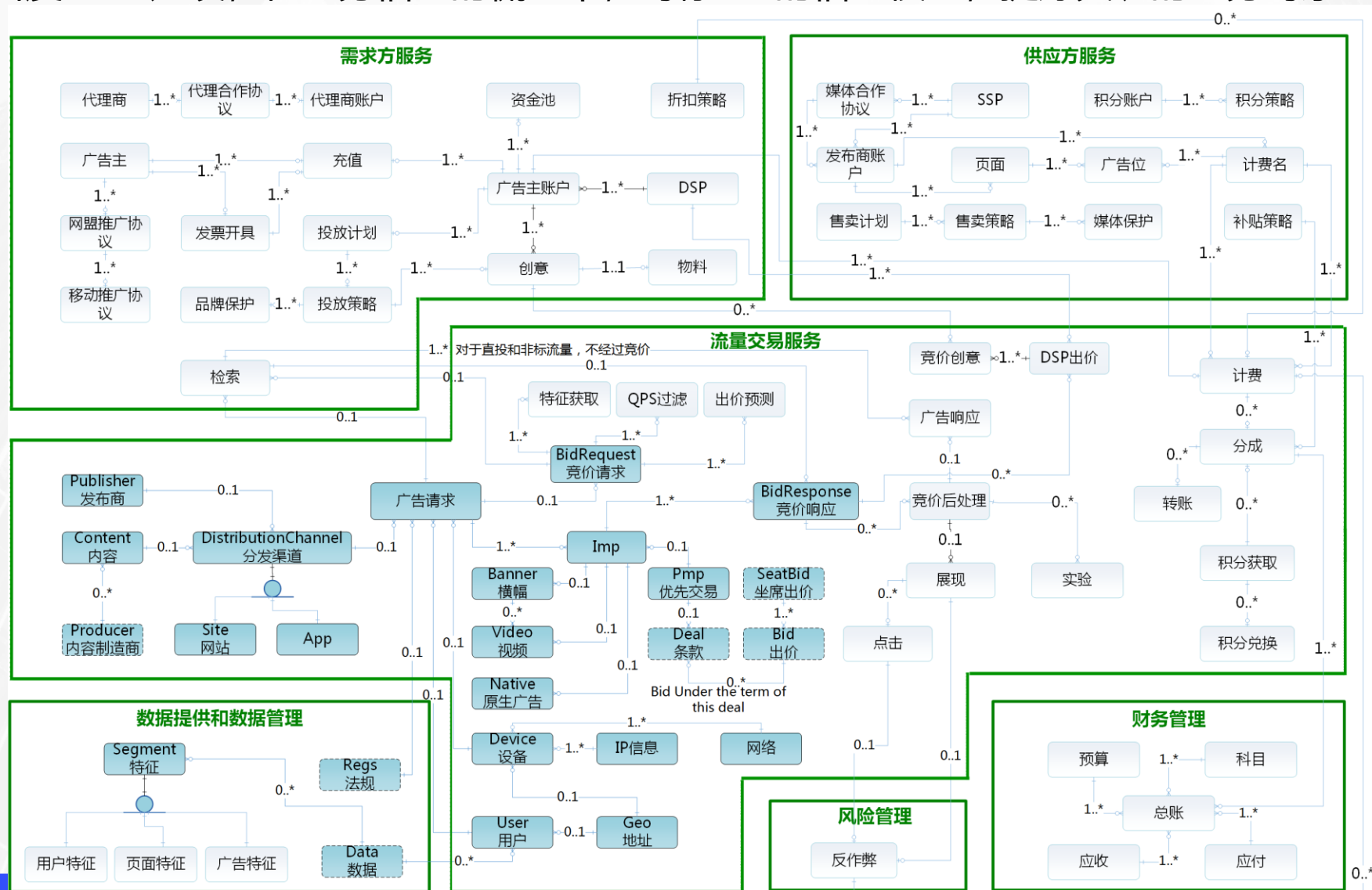
# 案例一：百度展示广告EDW建设

- 百度EDW建设，数据模型的设计出发点，来自于企业架构中的业务架构，也就是**业务信息**

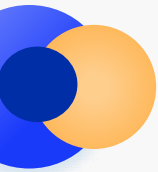


## 案例一：百度展示广告EDW建设

- 百度EDW建设，在业务信息的梳理中，对标IAB的信息模型，提炼关键的业务对象

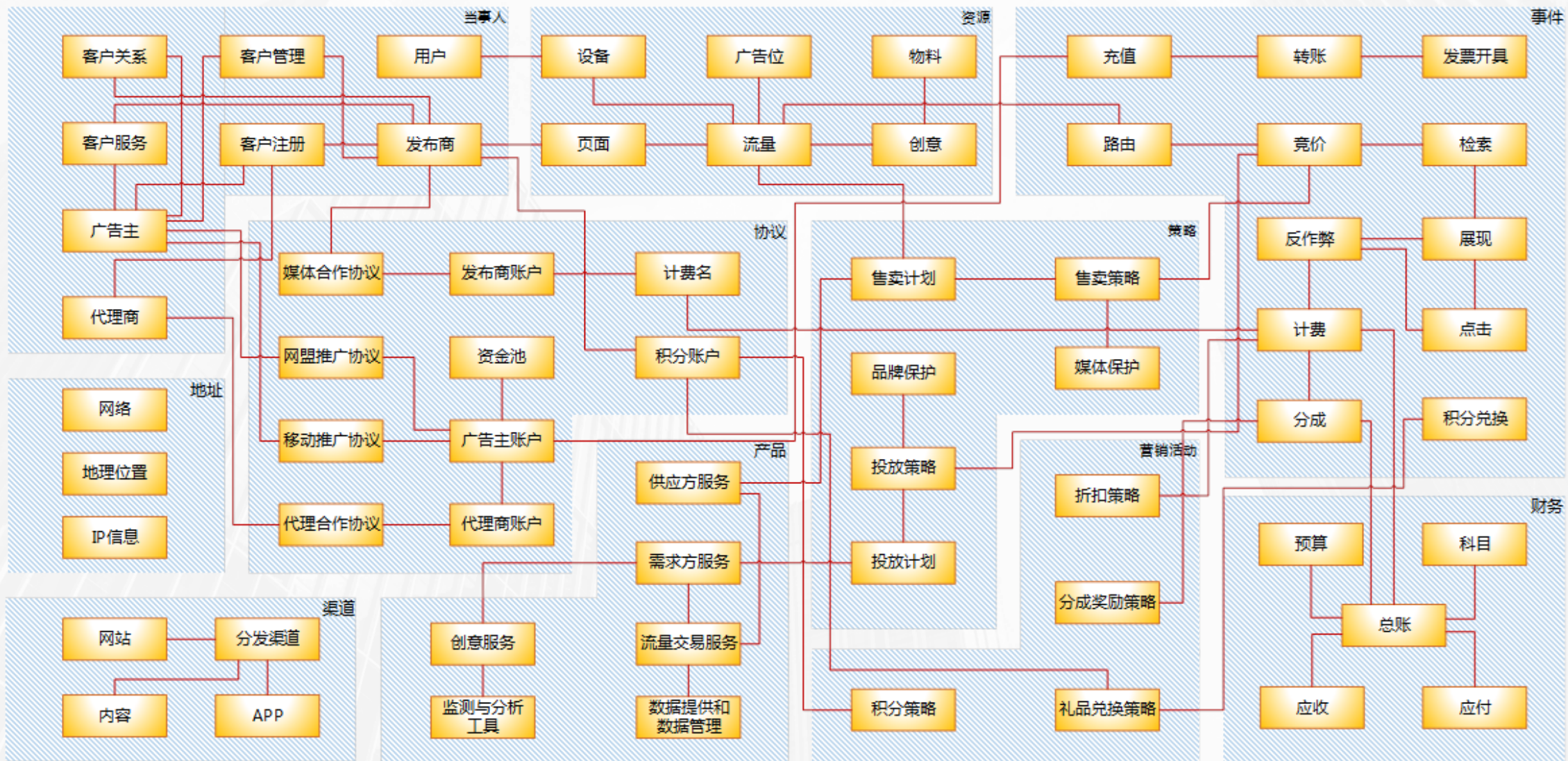




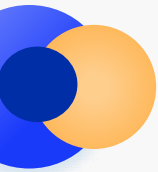


# 案例一：百度展示广告EDW的主题域划分

- 按照数据仓库设计的流程，我们划分了10主题域，梳理了关键数据实体。

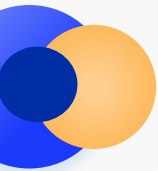






## 案例二：华为数据底座





# 案例三：银行数据仓库模型设计

## 金融数据仓库模型设计方式

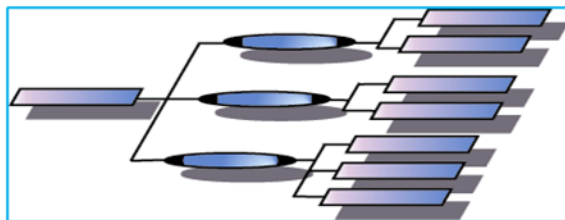
### 自顶向下

金融数据模型从**顶层规划**开始，基于业务能力（流程）→梳理业务对象→抽象数据实体→实体细化



### 全局覆盖

全面覆盖业务场景，不局限于某特定需求或场景

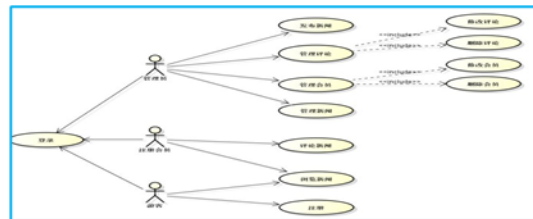


### 考虑未来

由于从架构层面梳理，考虑行业的通用性、未来一段时间内变革，能推动架构重构



## 传统数据模型设计方式

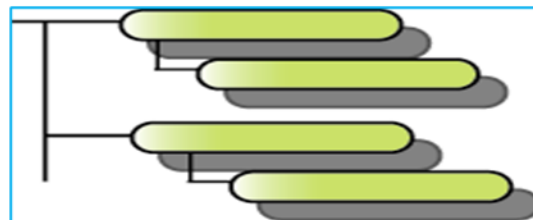


### 自底向上

传统数据模型基于功能或者数据需求，自底向上设计

### 局部覆盖

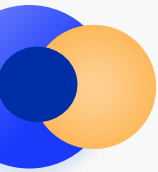
由于是需求驱动，只能特定需求或场景



### 基于现状

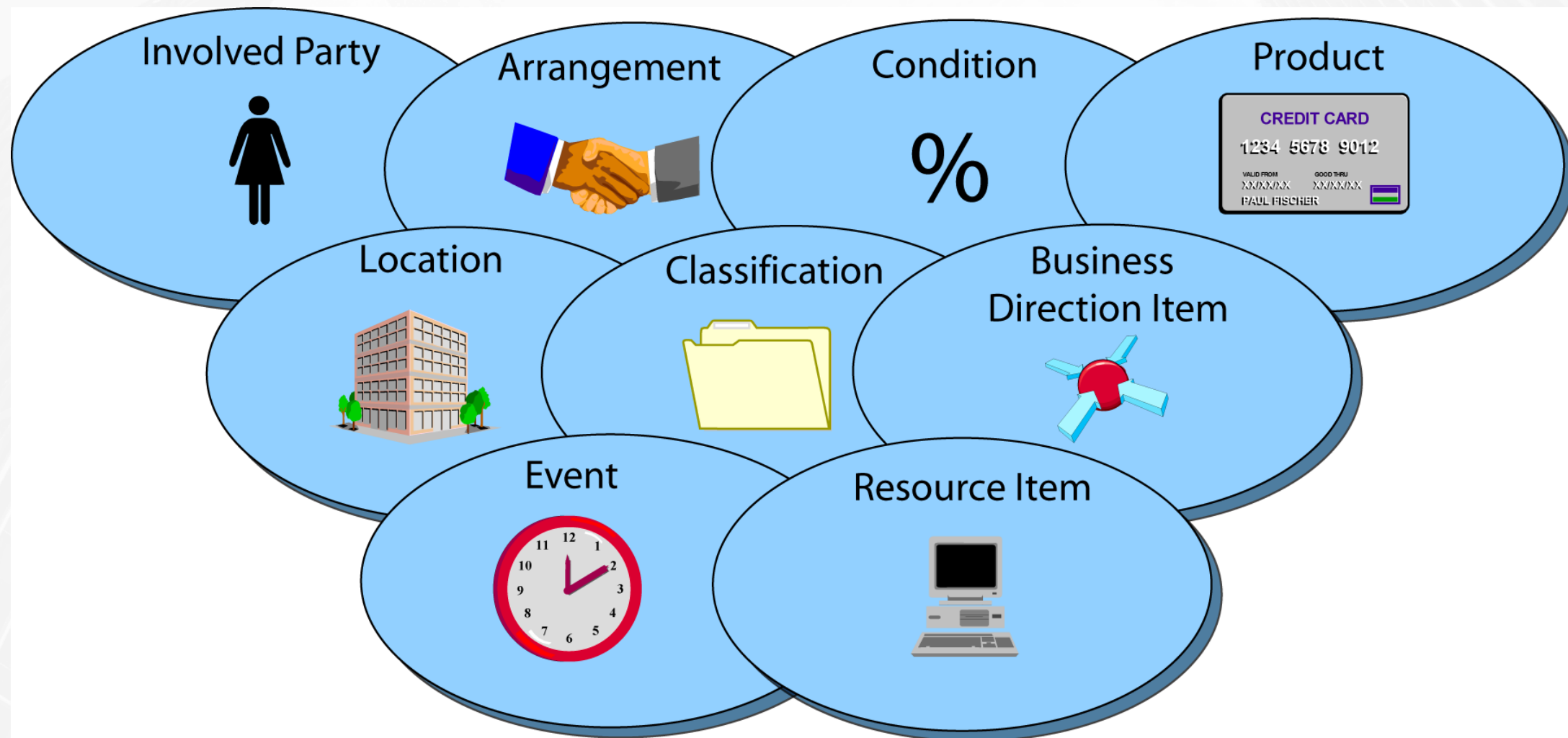
基于当前的业务流程、应用功能和数据现状，设计数据模型，很少考虑业务变革





# IBM数据仓库模型

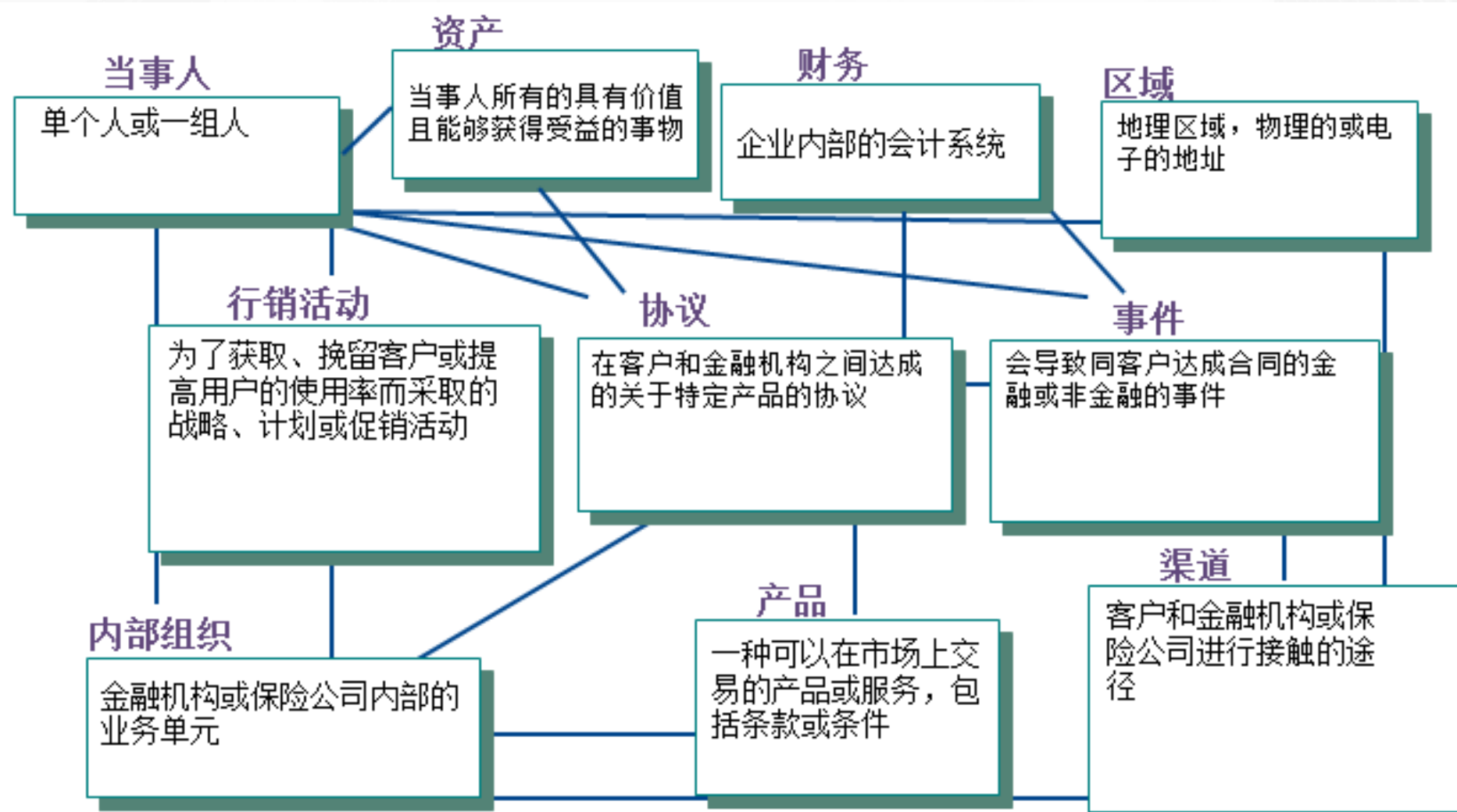
- 在银行业，**IBM BDM 9**大主题域包括：当事人（相关方）、产品、协议、事件、地址位置、分类及渠道、条件、资源项、业务方向。





# Teradata数据仓库模型

- 在银行业，Teradata FS-LDM 10大主题域包括：当事人、产品、协议、事件、区域、渠道、行销活动、资产、财务、内部机构。





# 04

## 心得总结







# 心得总结

- **顶层设计**：Onedata 体系建设，**需要顶层设计**，从业务架构推导出数据模型。
- **数据分层**：数据仓库中，需要不同的分层，来保证数据模型稳定性、可扩展性和可用性，以便数据模型适应业务场景的变化。
- **规范性**：数据模型的设计规范性、数据标准化（命名、值域等）、服务接口

— THANKS —

# 感谢您的观看

演讲人：邓成聪—独立顾问