

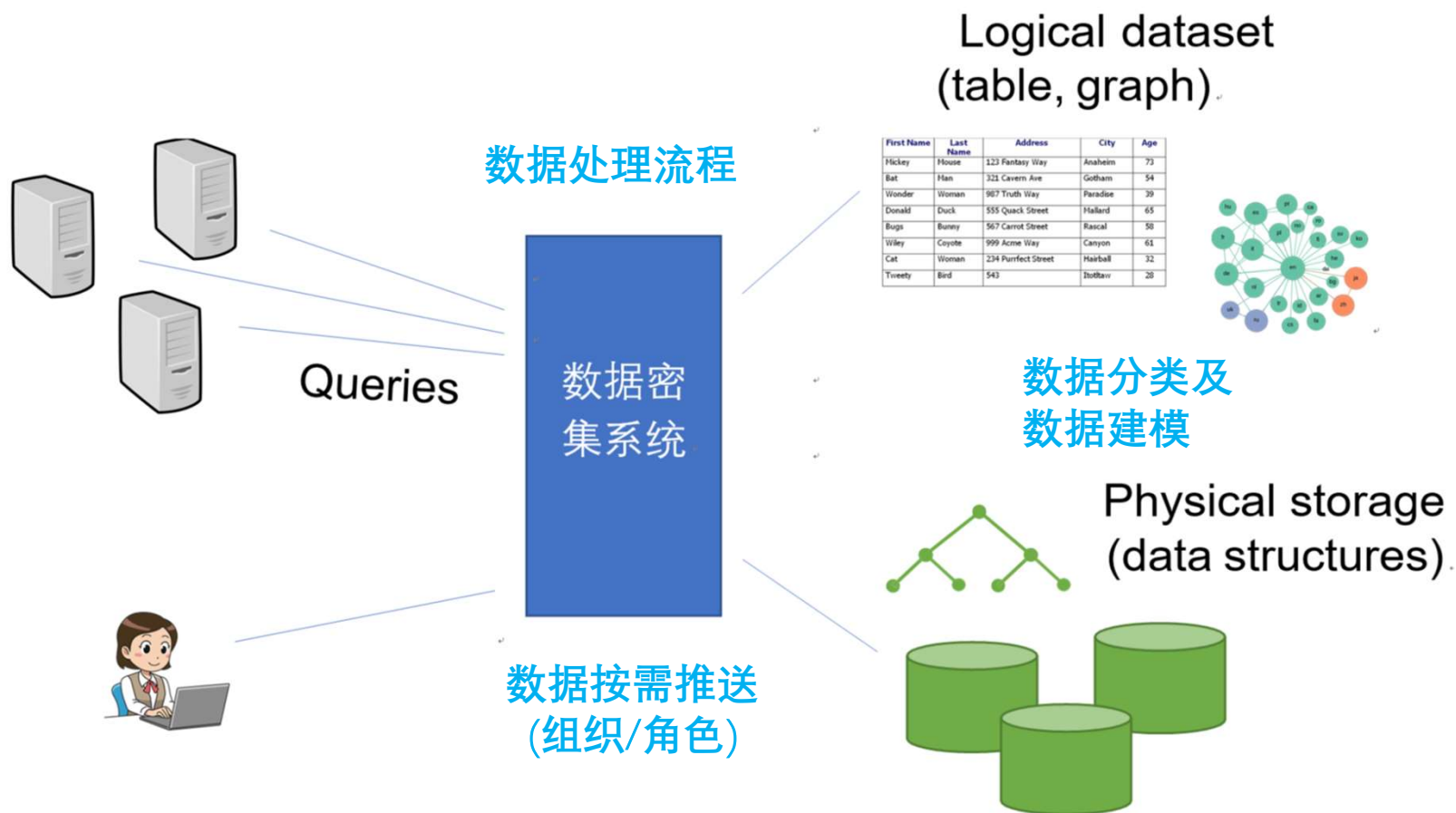


- 1 业务架构驱动的应用架构设计
  - 从模型到软件的转换-业务模型-建模方法发展
- 2 基于活动的过程建模方法
  - 业务任务规划-软件功能设计-活动时序流描述-活动执行控制
- 3 基于数据的过程建模方法
  - 数据分类-数据建模-数据流图-组织建模
- 4 基于状态的过程建模方法
  - DEDS-经典Petri网-高阶Petri网-PNG流程建模-PNG仿真实例
- 5 基于事件的过程建模方法
  - EPC-EPC规则语义-EPC建模规范-企业管理基础-ARIS实施实例
- 6 小结

### 3 基于数据的过程建模方法



数据建模，主要涉及到按照用途的数据**分类**及数据**建模**、数据动态处理（**数据流**），以及数据按需（**组织/角色**）推送等方面。

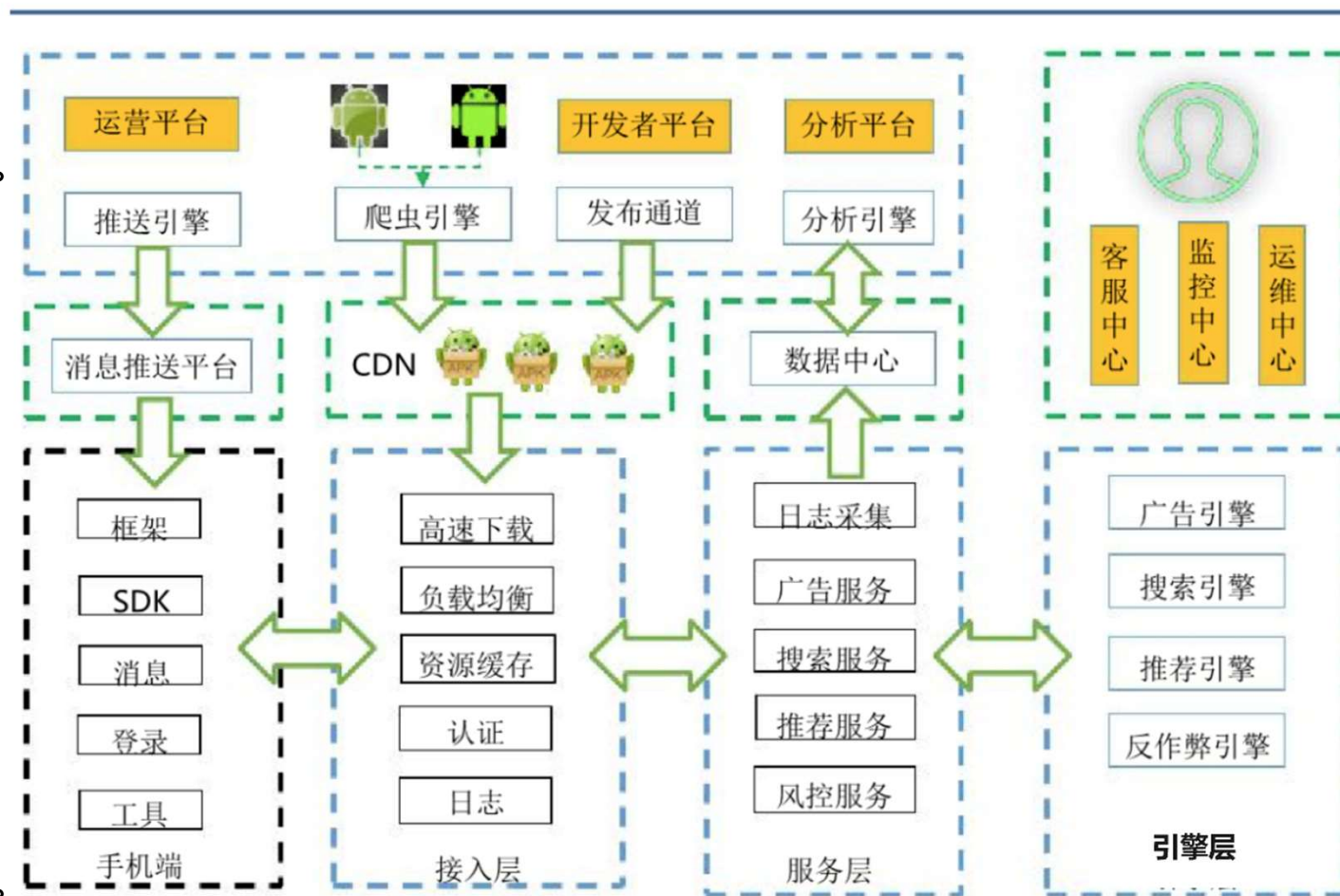


数据密集系统，例如内容分发系统等。

以数据操作为核心，主要特征包括：

- 高效数据I/O操作；
- 数据**分类**索引；
- 数据可扩充性存储；
- 数据多模态智能处理；
- 数据**按需**推送分发；

如何实现**数据分类**、如何支持数据按需服务，涉及到**数据建模**、**组织建模**。



数据类软件的系统架构



- 数据划分方面，而不同的企业也建立了自己的数据分类方式。
- SAP认为ERP等系统中的数据主要分为主数据和业务数据两大类。
  - 所谓主数据，就是指客户、产品、供应商等在相当长一段时间内不会被频繁创建或修改的“静态”信息；
  - 而业务数据指的是如订单、报修单等随时会被创建或处理的“动态”凭证。
- Oracle认为：
  - 主数据是描述核心业务实体的关键事实(例如客户、产品、员工、地区等)和这些事实之间的数据关系，这些关键事实在多个业务系统中被用到反复用到，在多个IT系统交互时，相同的实体可以有所区别。

➤ Microsoft认为:

- 主数据是改变缓慢，并可以跨系统共享的业务关键名词，通常包括人、物、地点和抽象对象等业务实体的信息。
- 企业的数据分为五类：主数据、交易数据、元数据、分级数据和非结构化数据。

➤ IBM认为:

- 主数据是用来描述企业核心业务实体的数据，比如客户、合作伙伴、员工、产品、物料单、账户等；它是具有高业务价值的、可以在企业内跨越各个业务部门被重复使用的数据，并且存在于多个异构的应用系统中。
- 主数据定义企业核心业务对象，主数据还包括关系数据，用以描述主数据之间的关系，如客户与产品的关系、产品与地域的关系、客户与客户的关系、产品与产品的关系等。
- 主数据的特点包括：它是准确的、集成的，其次它是跨业务部门的，再有就是它是在各个业务部门被重复使用的。
- 数据管理的内容及范畴通常包括交易数据、主数据以及元数据。

➤ 按照时间稳定性以及应用的特点，把数据分类为：

- **基础数据（主数据）**：主要是相对稳定、可以在公共范围内重用的基础数据，可独立存储并单独存在，主要的操作可以分为只读不写，以及单独的增删权限等。
- **业务数据**：是大量生产及重复处理的业务数据。
- **状态数据**：面向流程类应用的信息控制所建立的数据，主要体现为主数据或者事务数据的动态状态变化。
- **元数据**：数据管理相关的数据。
- **非结构化数据**：模型、视频、语音等多模态数据，在大数据时代越来越多。





- 针对不同的数据用途，如何组织数据，更好地满足使用需求。这里就涉及到数据建模问题。
- 数据建模（Data Modeling）
  - 是对现实物理世界的事物的各维度数据进行抽象组织，其过程是一种用于定义和分析数据的要求，以及相应支持系统的过程。
  - 从抽象程度，企业数据可分为主题域模型、概念模型、逻辑模型、物理模型。具体包括：
    - 主题域模型：更多是管理方面的考虑。
    - **概念建模**：建立实体、实体的特征以及它们之间的关系。
    - **逻辑建模**：概念模型具体化。要实现概念模型所描述的东西，需要那些具体的功能和处理那些具体的信息。
    - **物理建模**：物理模型就是针对上述逻辑模型的内容，在具体的物理介质上实现出来。



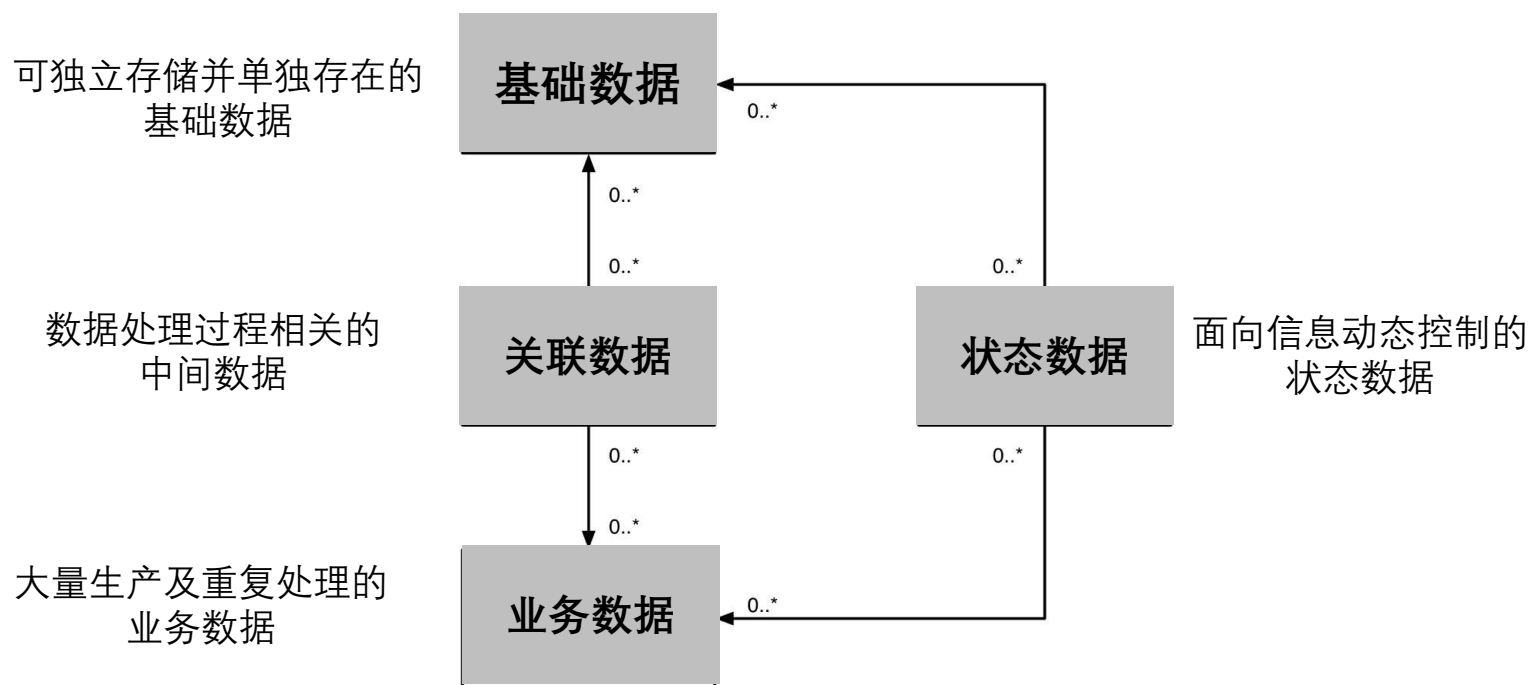
- 从关系数据库来说，数据表可以分为1-5NF，第五范式则是最高要求
- 第一范式（1NF）：业务属性不可再分（表中的每一列都是不可拆分的原子项）；
  - 第二范式（2NF）：满足1NF基础上，并且没有部分依赖。（非主属性都要由主属性来唯一确定）
  - 第三范式（3NF）：满足2NF基础上，没有传递依赖。或者说，关系重复，能互相推导出来。
  - BC范式（BCNF）
  - 第四范式（4NF）
  - 第五范式（5NF）



➤ 数据几乎总是用于两种目的：

- 操作型数据：一般仅反映数据的最新状态，按单条记录事务性来处理；其优化的核心是更快地处理事务。
  - 分析型数据：往往是反映数据一段时间的状态变化，按大批量方式处理数据；其核心是高性能、多维度处理数据。
- 
- 通常操作型系统简称为OLTP（On-Line Transaction Processing）联机事务处理
  - 分析型系统简称为OLAP（On-Line Analytical Processing）联机分析处理。

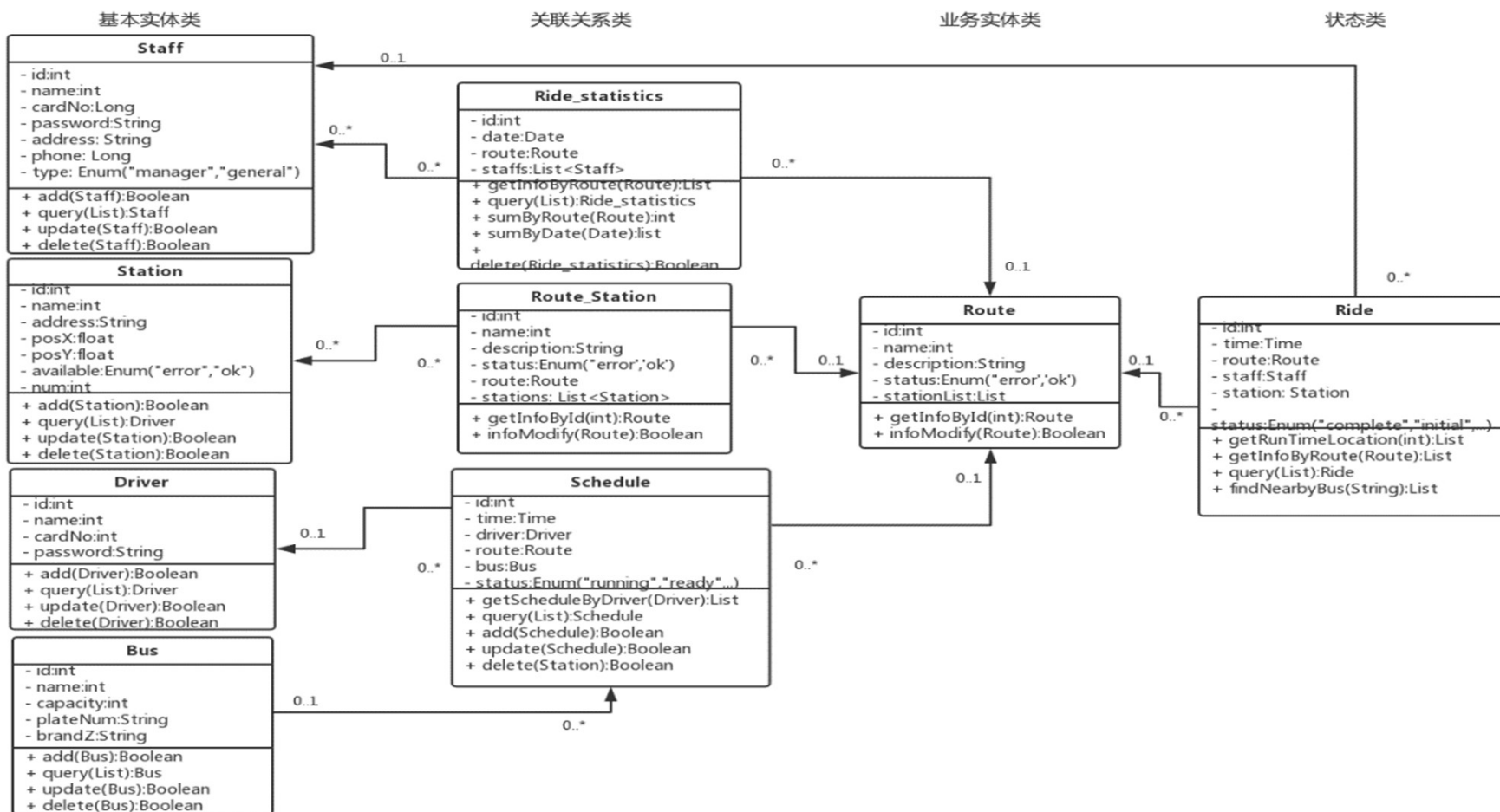
## ➤ 数据建模方式往往涉及到各类数据的建模和关联：



# 一个数据建模实例



- 乘客、站点、司机和车辆被抽象为一些基本实体，而面向运行过程中的调度、计数、路线等抽象为业务实体或关联实体。



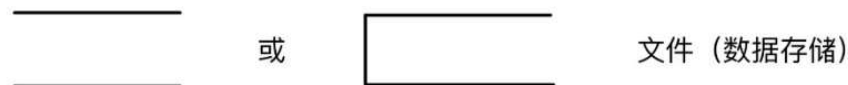
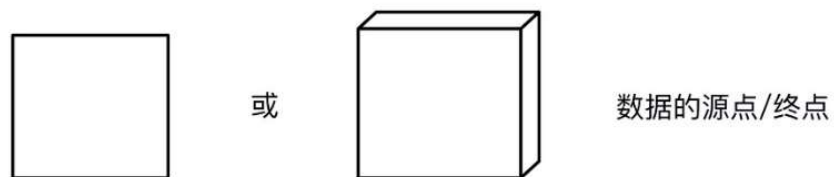
## 3.3 数据流描述



➤ 数据流图(DFD): 它从数据传递和加工角度, 以图形方式来表达系统的**逻辑功能**、数据在系统内部的逻辑流向和逻辑变换过程。

➤ 数据流图的主要元素:

- 外部实体 (数据源或数据终点)
- 处理活动
- 数据存储
- 数据流连接



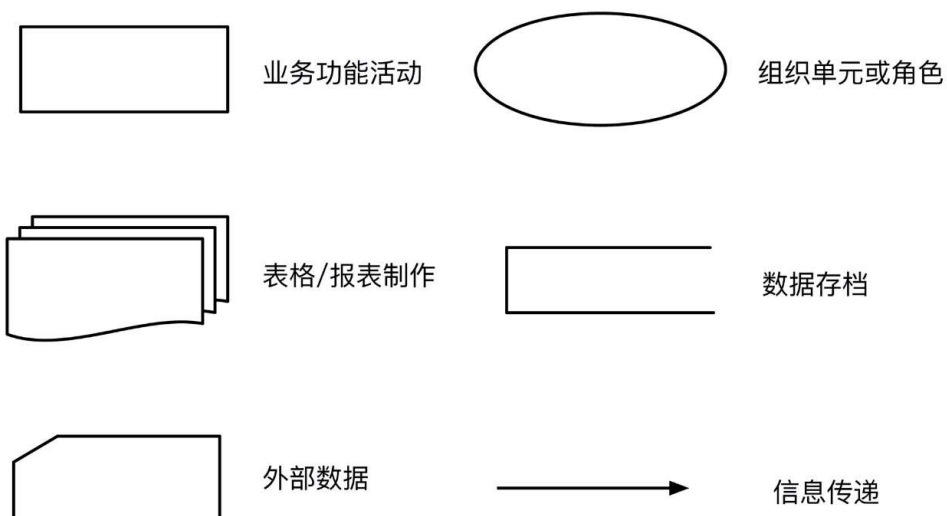


➤ 数据流图绘制是，需要注意以下原则：

- 所有的数据流必须**以一个外部实体开始**，并以一个外部实体**结束**。
- 外部实体之间**不应该存在数据流**。
- **保持数据守恒**。一个（处理）活动所有输出数据流中的数据必须能从该加工的输入数据流中直接获得，或者说是通过该加工能产生的数据。
- 每个（处理）活动必须**既有输入数据流，又有输出数据流**。
- 一个（处理）活动的输出数据流不应与输入数据流同名。
- 分层的DFD图描述复杂的业务逻辑。
- 分层DFD图可以分成：顶层DFD、中层DFD和底层DFD。
- 底层DFD是指其活动不能再分解的数据流图



- 业务流程图(TFD)是结构化分析方法中的**系统分析**模型,它用一些尽可能少的规定符号及连线来表示某个具体业务处理过程。
- 业务流程图的基本符号



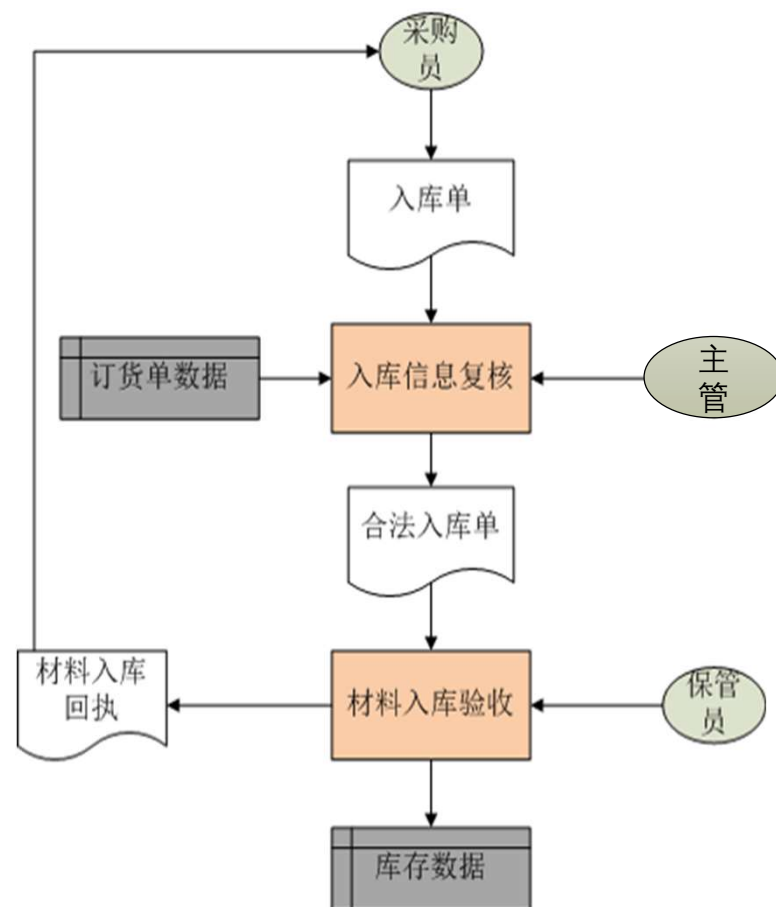
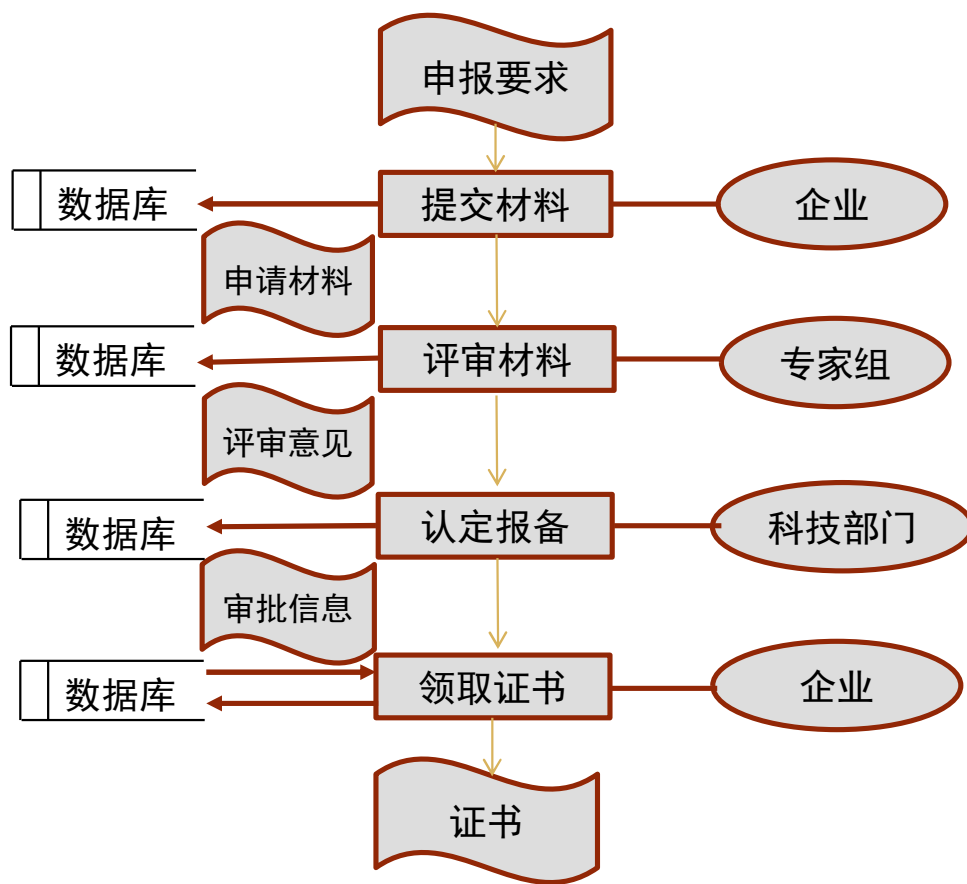


## 业务流程图



- 业务流程图：描述处理活动的**关系**，业务活动与组织单元的**对应关系**，以及业务活动与数据读写**存储**，以及外部的**信息流/文档流**。

➤ TFD变形

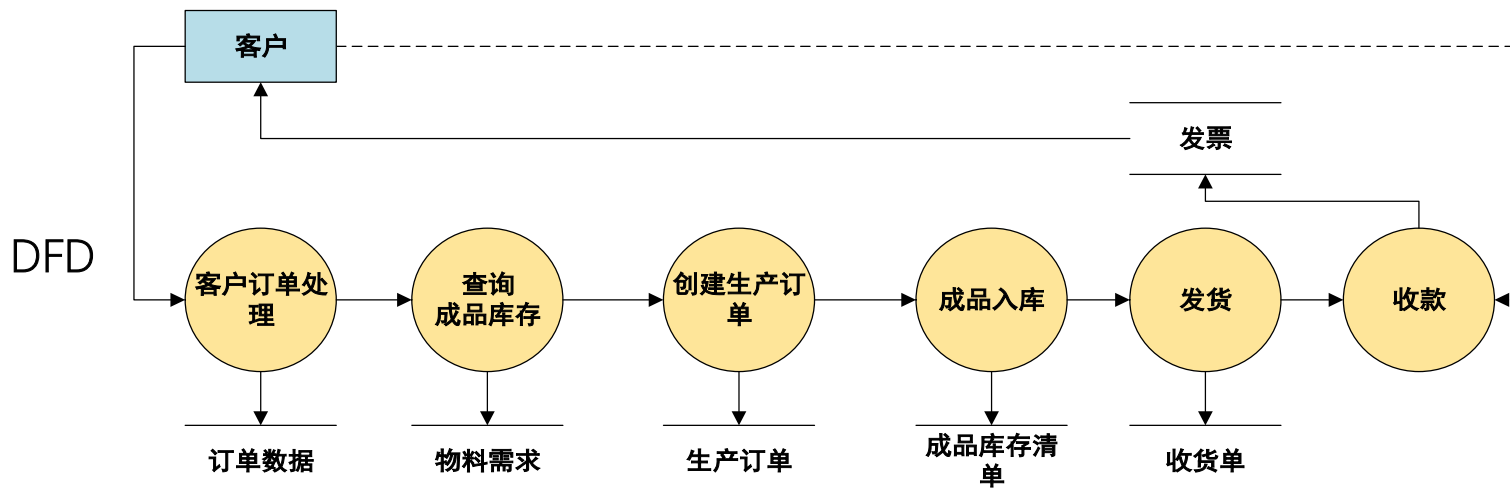
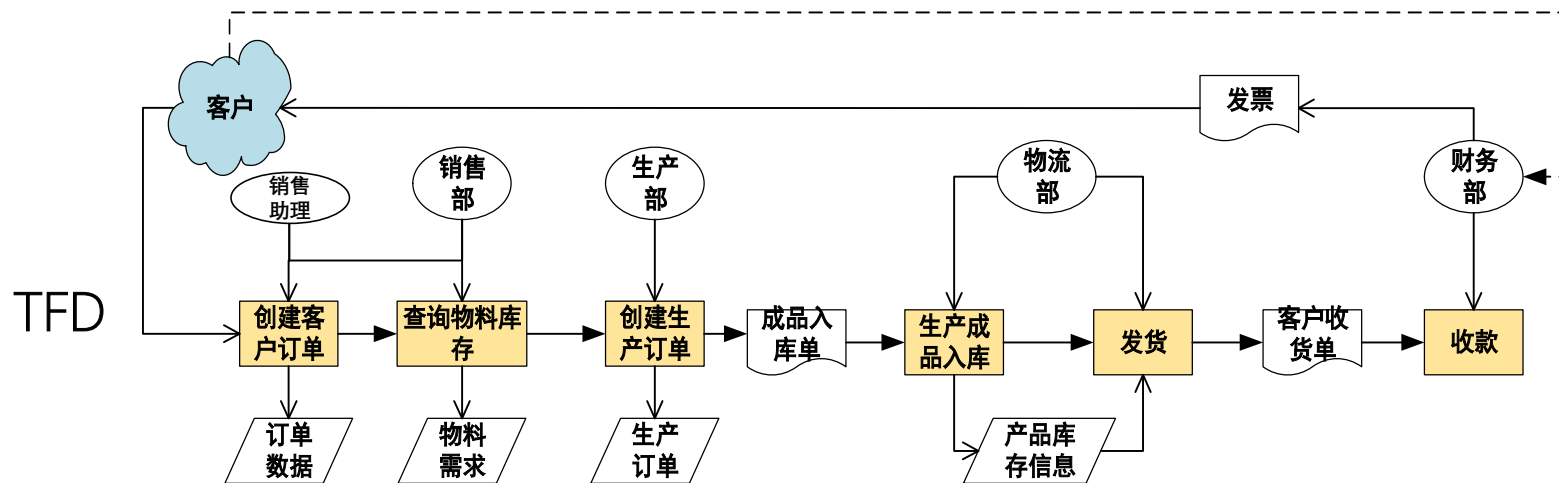




➤ 业务流程图TFD和数据流图DFD 的关系如下：

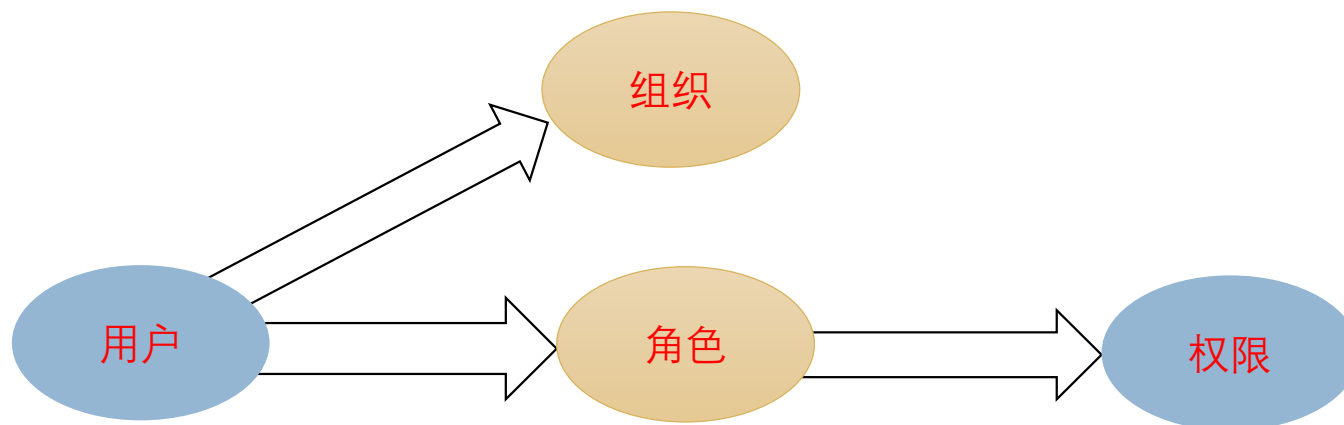
- TFD 和 DFD 都是描述业务数据处理过程的**图形建模**方法。
- 从描述的重点来看，TFD 强调业务过程中每一项处理**活动**和具体**业务**部门的关系。而 DFD 更注重描述业务内**数据间的关系**及业务的“系统”特征，标识业务通过外部实体与其环境交换信息。
- 从使用者的角度来看，用 TFD 描述企业各项业务的数据处理过程更容易与用户进行交流。DFD 较 TFD 抽象，描述的是业务处理过程的数据处理模式，但难于描述系统的控制流。

# 业务流程图和数据流图的对比





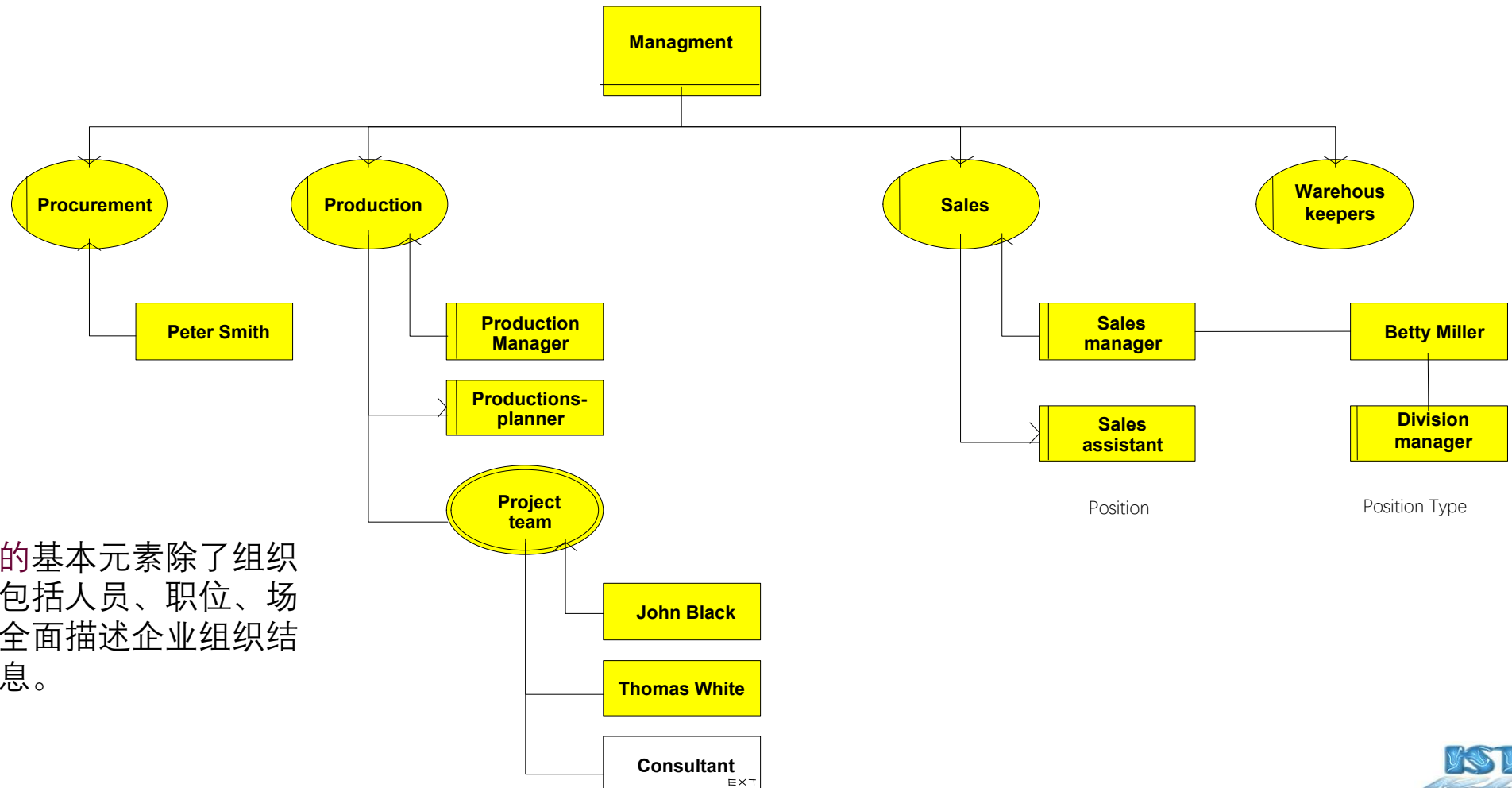
- 每一种软件和开发这个软件的团队组织是对应的。
- 数据的应用中涉及到面向特定用户的**权限控制**，**信息推送**等，因此，组织模型构建也是静态建模中的基础内容。
- 权限控制主要有RBAC（Role-Based Access Control）基于角色的访问控制。
- RBAC认为权限的过程可以抽象概括为：Who是否可以What进行How的访问操作。
- 这个逻辑表达式将权限问题转换为Who、What、How的问题。who、what、how构成了访问权限三元组。



## 组织机构图示例



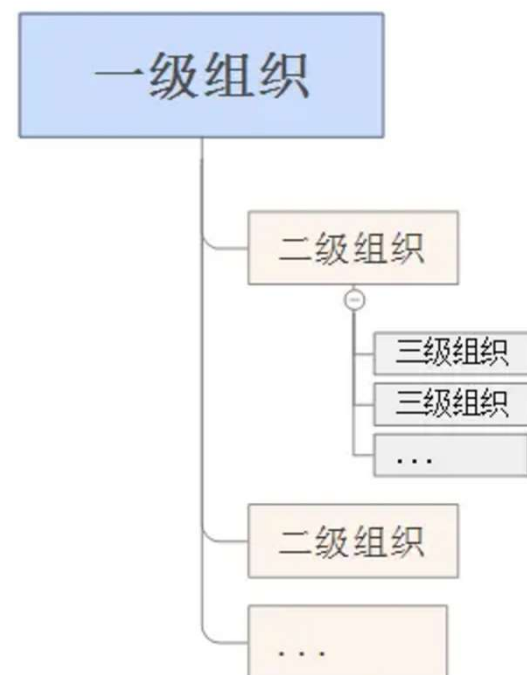
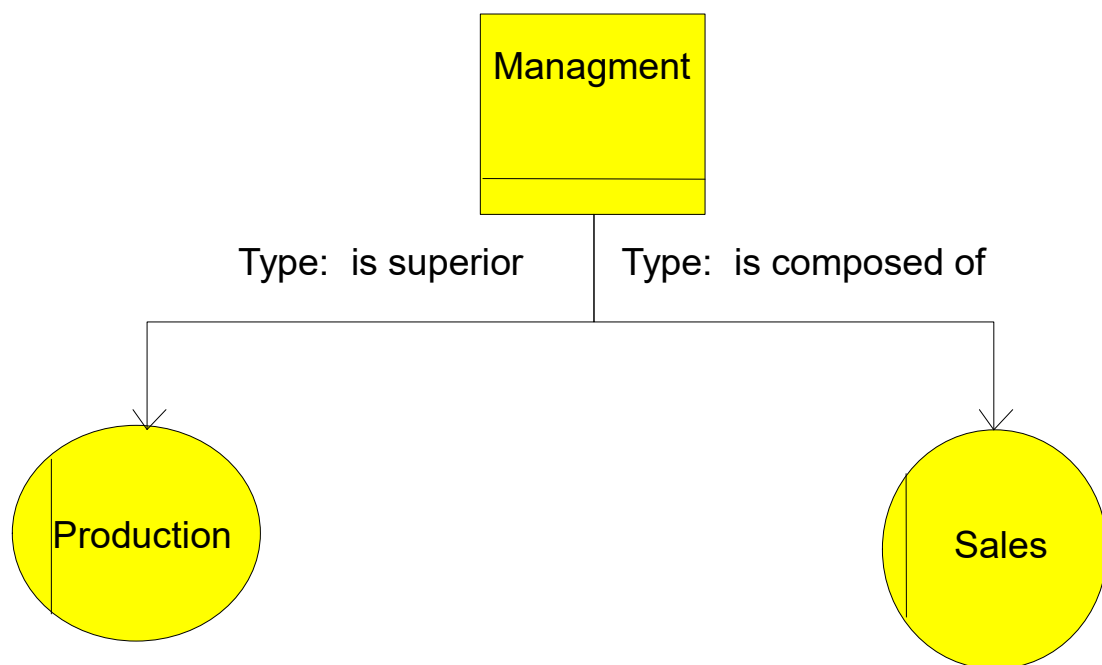
- 组织视图描述了公司任务操作者之间的静态关系，使用组织机构图(organizational chart)的方式描述组织单元、组织的分解关系、层次结构、隶属关系等。



组织机构图的基本元素除了组织单元外，还包括人员、职位、场所等图元，全面描述企业组织结构的各种信息。

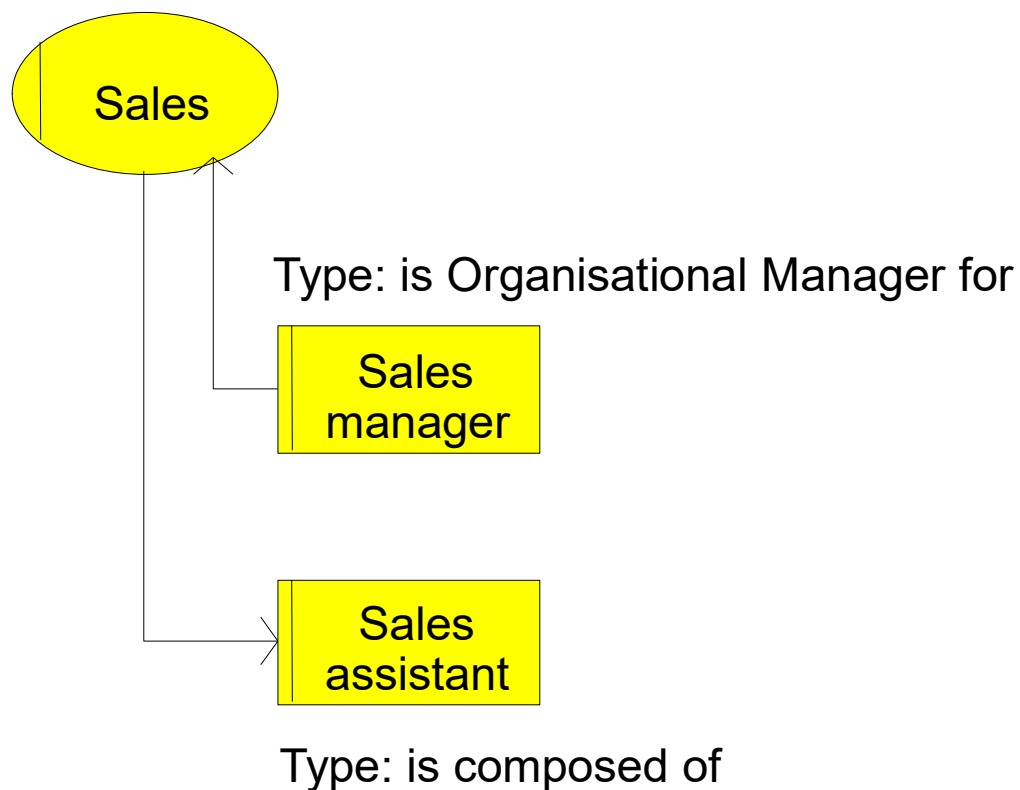


## 数据可视的层化化权限

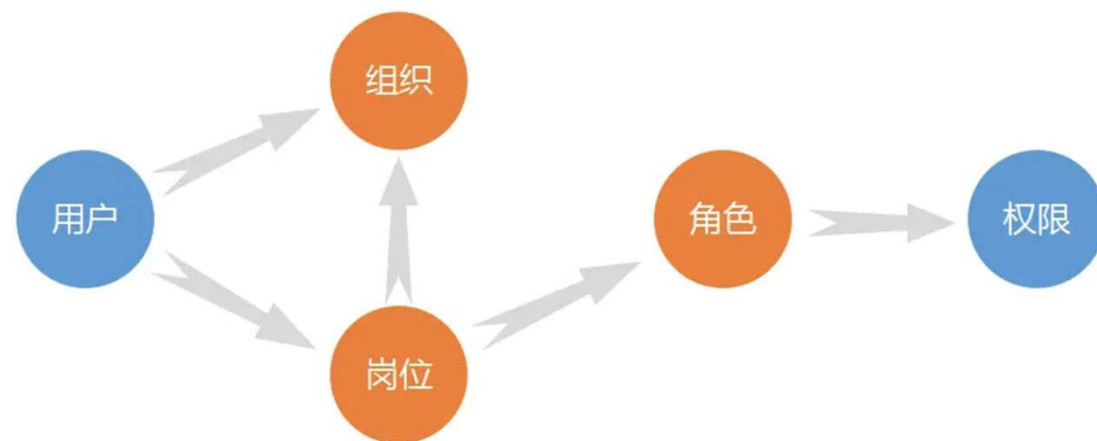




- Position（职位）是公司内最小的组织要素。相关的工作职责和管理权限定义在相应的描述中。

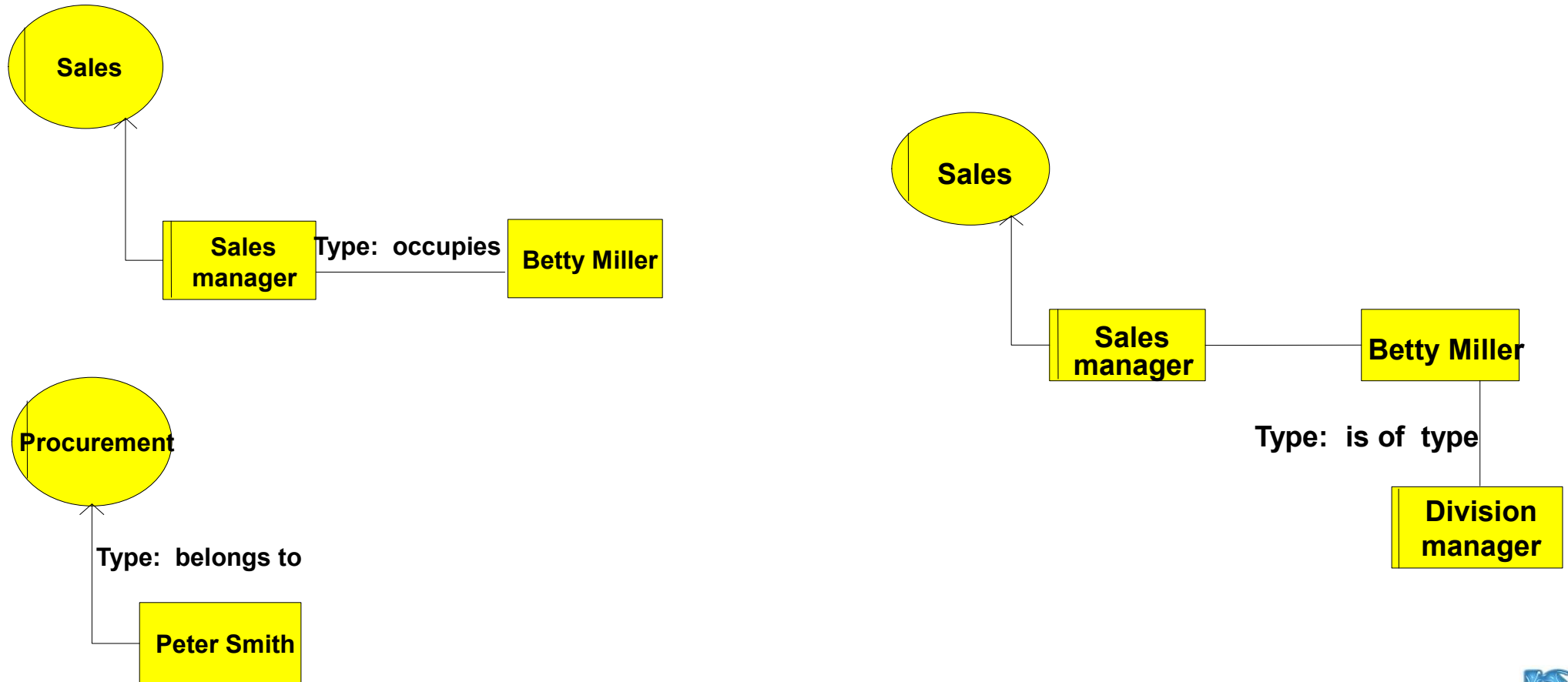


组织和岗位、岗位和角色提前进行关联



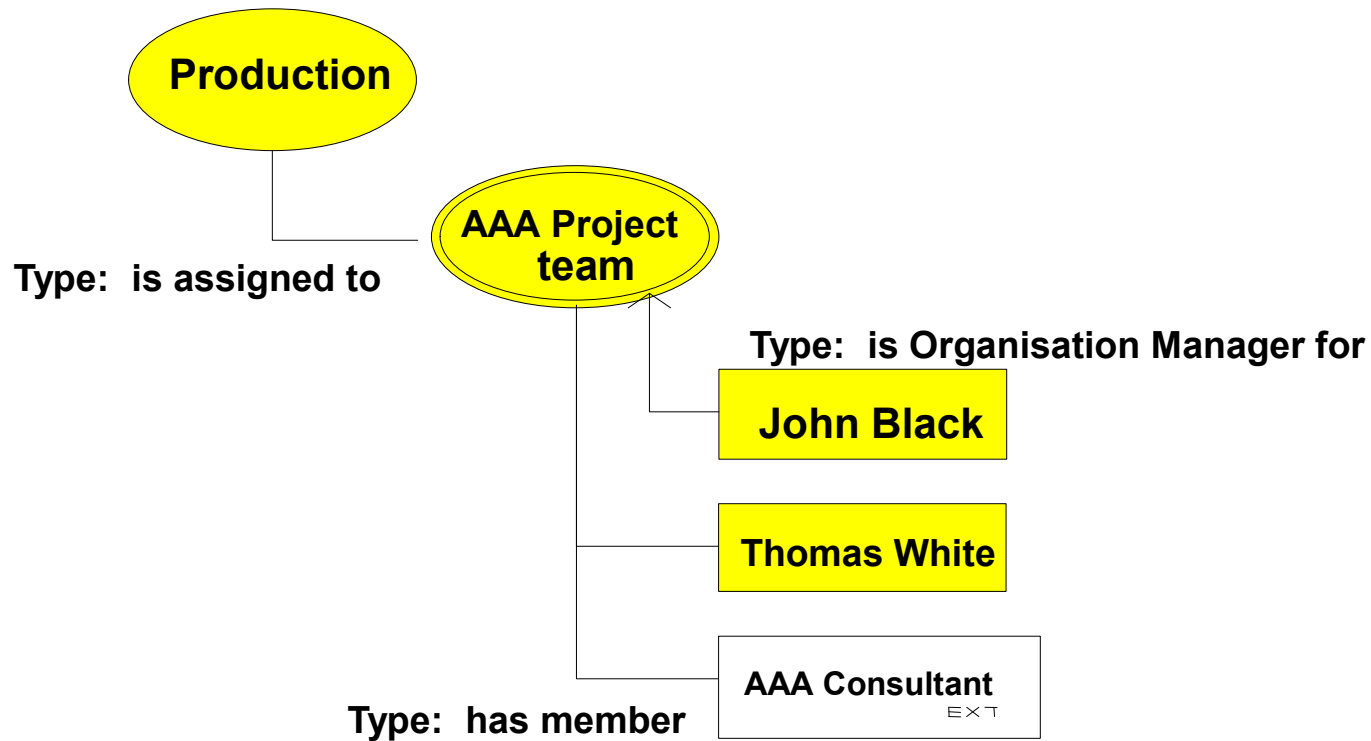


- Persons (人员) 是被赋予独特的人事代码的公司的具体雇员。
- Person type (人员类型) 代表具有相同性质的单一雇员 (即权利或责任)





- Group（组）代表一组员工/人员，他们在有限的时间内一起完成特定的任务（例如项目团队）。
- Member（成员）可以指向对应人员，也可以设置虚拟成员对应外部团队。



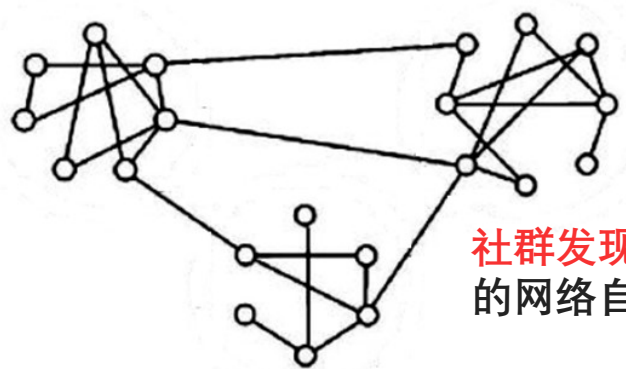


## ➤ 大企业（集团内或子公司间）的组织形式：

- U型组织结构：按职能划分成若干部门，即企业实行集中控制和统一指挥。
- M型组织结构：多事业部相对集中管理，便于协调和控制。
- H型组织结构：多个法人实体集合的母子体制，主要靠产权纽带来连接，各子公司保持了较大的独立性。

## ➤ 企业间组织形式：

- 联合体
- 链式（产业链上下游）
- 网式自组织



社群发现 (Community Detection)  
的网络自组织



➤ 在智能互联时代，确定合适的用户及用户群，协助对应用户群找到最适合的信息，如何破除“信息茧房”成为了一个重要的出发点。

➤ 用户画像（Persona）：

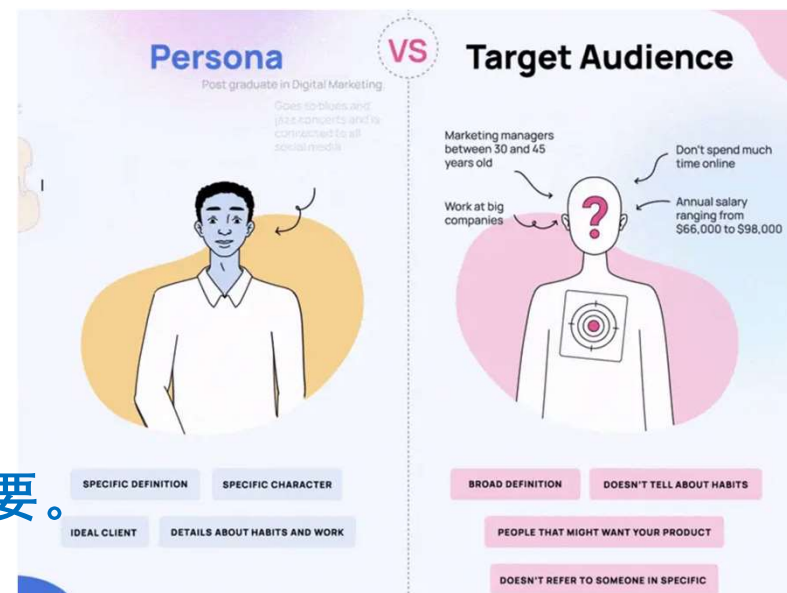
- 用户画像的构建和应用都基于深入的数据分析和对**用户群行为**的精准把握。
- 通过不断优化用户画像，企业能够更好地满足用户需求，实现商业价值的最大化。

➤ 信息茧房

来源于美国法学教授凯斯·桑斯坦的《信息乌托邦》

（2006），是指当个体只关注自我选择的或能够愉悦自身的内容，而减少对其他信息的接触，久而久之，便会像蚕一样逐渐桎梏于自我编织的“茧房”之中。

➤ 因此，根据主体或用户群，选择合适推荐算法便非常重要。



# 信息推荐主要算法



算法分类	子分类	子子分类	思想	推荐步骤
协同过滤算法	基于邻域的方法 neighborhood-based	基于用户的协同过滤算法 UserCF	给用户推荐和他兴趣相似的其他用户喜欢的物品	1. 找到和目标用户兴趣相似的用户集合 2. 找到这个集合中的用户喜欢的, 且目标用户没有听说过的物品推荐给目标用户
		基于物品的协同过滤算法 ItemCF	给用户推荐和他之前喜欢的物品相似的物品	1. 计算物品的相似度 2. 根据物品的相似度和用户的历史行为给用户生成推荐列表
	隐语义模型 latent factor model	/	通过潜在特征联系用户和物品	1. 基于用户的行为统计, 按用户兴趣给物品分类 2. 确定用户对哪些类别的物品感兴趣, 以及感兴趣的程度 3. 确定不同类别物品的权重, 选择性推荐给用户
	基于图的随机游走算法 random walk on graph	/	/	1. 将用户和对应的行为用图模型表示出来 2. 在图模型中, 从某用户的对应节点开始进行随机游走, 最终得出物品的权重, 进行推荐
	其他	/	/	/
基于内容推荐	/	/	根据用户过去的浏览记录, 来向用户推荐用户没有接触过的推荐项	<p>主要问题:</p> <p>数据隐私问题 (Data Privacy Issues)</p> <p>推荐的多样性 (Diversity of Recommendations)</p> <p>冷启动问题 (Cold Start Problem)</p> <p>算法的复杂性 (Complexity of Algorithms)</p>
基于关联规则推荐 Association Rule-based Recommendation	/	/	在一个交易数据库中统计购买了商品集X的交易中有多大比例的交易同时购买了商品集Y	
基于效用推荐 Utility-based Recommendation	/	/	理论基础: 效用函数	
基于知识推荐 Knowledge-based Recommendation	/	/	建议某一领域的一套规则, 进行推荐	