

**导读：**在数字化转型的浪潮下，数据架构获得了越来越多的关注。作为企业架构中的关键纽带，数据架构解决了业务与数据间的映射，规范了应用架构中的数据集成关系，指导了技术架构的技术选型，在企业中发挥着不可或缺的作用。伴随DataOps等场景的出现，数据架构会越来越走向数据消费端，为我们带来更多的变化和新发现。

# 1、什么是数据架构？

自数据库技术兴起以来，人们慢慢发现需要一种东西去描述数据的状态、定义数据需求、指导数据集成、管理数据资产，以更好地实现数据策略——数据架构应运而生。

**数据架构**是识别企业数据需求，并以这些数据需求为出发点，设计和维护的主蓝图，以使用这些主蓝图来指导数据集成、控制数据资产，并使数据投资与业务战略保持一致（DAMA-DMBOK2，DAMA International）。

越是专业的定义越让人难以理解，举个例子：

小明打算在有生之年买一套北京二环内的房子，相关机构需要分析小明是否有买房资格，这是**业务需求**；那么为了分析小明的买房资格，最起码需要小明的身份数据、婚姻数据、在京纳税数据、在京缴纳社保数据、在京房产数据，这些是**数据需求**；为了有效获取并集成这些数据，以实现相关机构分析小明买房资格的业务需求，我们需要下图所示的**数据架构**（注：此处只是简单例子，企业真实数据架构含有很多组件，会复杂很多）。

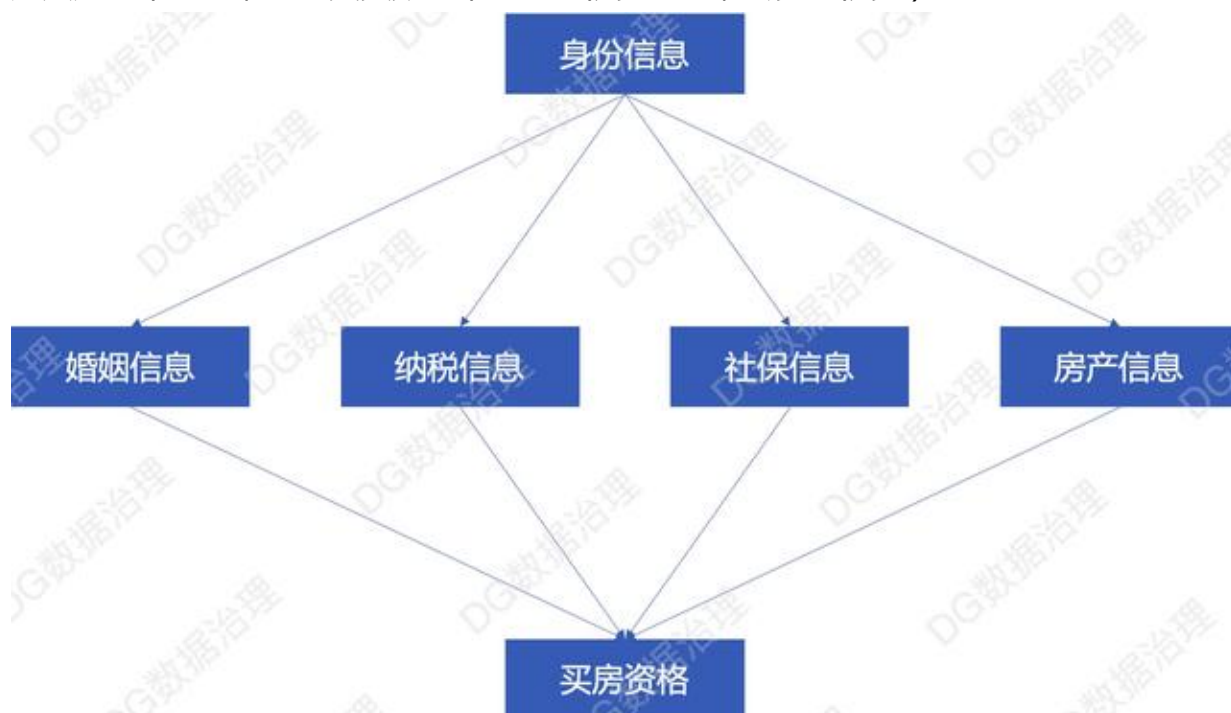
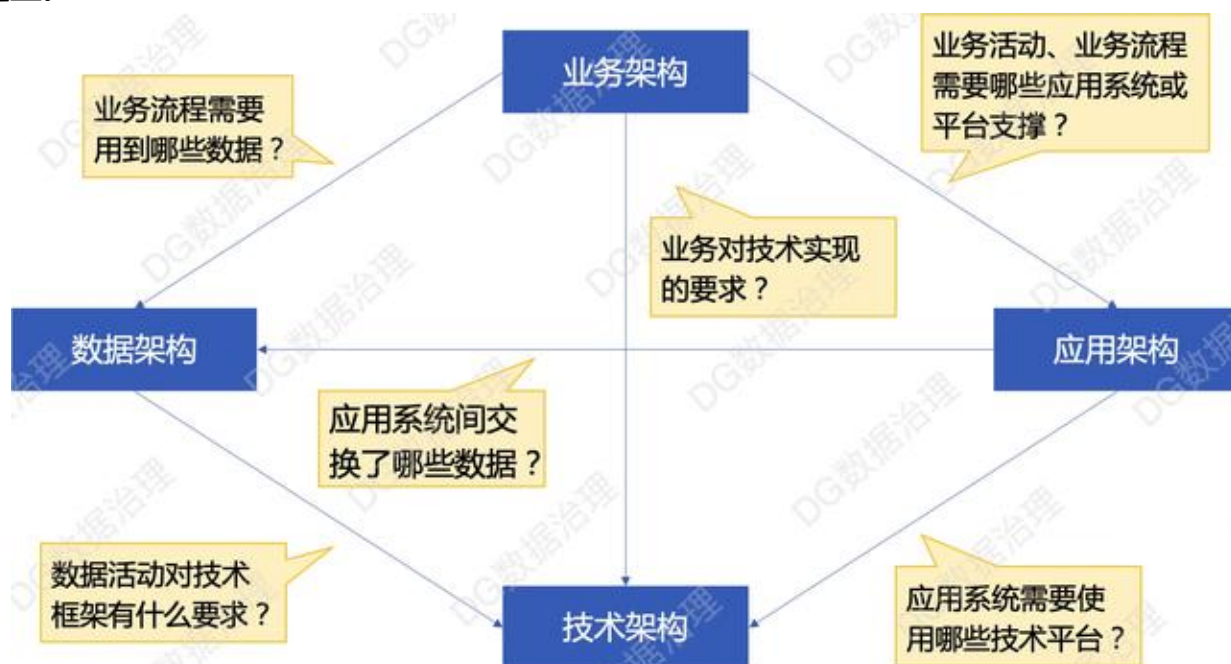


图 1分析买房资格的数据架构

## 2、 数据架构解决什么问题？

作为连接企业架构中其他三个架构（业务架构、应用架构、技术架构）的关键纽带，数据架构解决了业务与数据间的映射，规范了应用架构中的数据集成关系，指导了技术架构的技术选型。



数据中台盛行，DataOps兴起，数据架构才是未来

图2数据架构是企业架构中的纽带

具体来说，数据架构在企业中的作用，可总结为以下四点：

### 1、 从不同层次描述数据，为数据管理奠定基础

作为企业数据层面的主蓝图，数据架构管理了整个企业的复杂数据和信息交付，数字化时代，企业拥有的数据量远超出员工的理解范围，数据架构使企业能在不同的抽象层次上表达数据，为数据管理奠定了坚实的基础。

### 2、 定义企业数据状态，表达战略数据需求

数据架构是对企业当前数据状态的整体性描述，通过数据架构，企业数据状态一目了然，通过分析数据架构中的不足与业界的差距，可以帮助企业定义未来的状态目标，帮助企业快速

制定战略数据需求。

### 3、 促进企业数据标准化，指导企业数据集成

数据架构为企业提供了统一的数据语言（包括标准业务词汇表、标准技术元数据等），将业务架构中的业务需求转化成为了数据、系统、技术需求，使不同部门和不同身份的技术和业务人员在数据层面达成一致，为数据标准化和数据集成提供了良好的支撑。

### 4、 规范企业数据流转，拉通数据生命周期

数据生命周期中包括了数据的创建、获取、移动、转换、存储、维护、共享、使用、处理等诸多流程，在整个生命周期中数据可以被清理、转换、合并、更新和汇总，涉及到太多的数据流过程，规范数据在诸多生命周期阶段的流转细节可能会非常复杂，要求不同阶段之间相互理解并随时达成一致，此时数据架构必不可少。

## 3、 数据架构包含哪些内容？

作为企业层面的"主蓝图"，数据架构中包含诸多组件，企业需求的不同可能导致其数据架构包含的组件也不相同，以下是数据架构中的一些常见组件：

### 1、 数据模型——数据架构的核心

· **概念数据模型**：概念数据模型是能表示现实世界的概念化结构，通过概念实体及关系，从业务的角度对信息进行的高层级描述。如在文章最开始的例子中，小明买房的整体建模就可以理解为是一种简单的概念模型，概念模型是为了让业务能明白数据有什么，帮助数据与业务沟通。

· **逻辑数据模型**：逻辑数据模型在概念数据模型的基础上定义了各个实体的属性，是对概念模型的进一步细化，包括所有的实体、实体的属性、实体之间的关系以及每个实体的主键、实体的外键等。逻辑模型帮助数据与应用沟通，比如应用需要小明的身份信息，逻辑模型里就需要有个身份信息的数据实体，并且需要明确身份信息中的各个属性（如身份证号、姓名、年龄等）。

· **物理数据模型**：物理数据模型是在逻辑数据模型的基础上，综合考虑各种存储条件的限制，将逻辑数据模型中的实体、属性以及关系转换成的物理元素（表、字段、索引等）。物理数据模型是帮助数据与技术沟通，明确了数据到底怎么建，存储的位置等。比如小明的身份信息具体存在那个物理表中就是在物理模型中明确的。（注：物理模型虽然属于数据架构的组件，但不是数据架构的产物）

## 2、数据流——数据与数据间的关系

数据流中主要需要设计数据如何流动，如小明需要买房，小明的社保信息需要从社保中心获取，小明的纳税信息需要从税务局获取，这些数据需要从社保中心、税务局流动到买房资格那里供分析使用，这就是数据流。数据流是数据架构很关键的部分，表示了数据从哪个源头来，到哪里去使用的过程。如身份信息就需要从公安获取，而不能从社保中心获取。数据流的设计也是数据架构设计的必备设计。

## 3、价值流——数据与业务间的关系

价值流更容易理解，所有的业务都有业务流程，而业务流程中会产生/使用/销毁数据，这里需要设计，哪些数据会在关键价值流中使用，对业务流程有关键作用，比如小明买房资格中的所需关键信息就是价值流需要设计的部分。

## 4、数据定义和数据标准——数据架构建设的准则和规范

在数据架构中需要规定数据模型的标准规范，包括标准的数据模型定义方式、模型属性的标准规范等，我们通常这个标准规范为数据标准。从理论上讲，所有的数据模型、数据流、数据映射关系的建设都需要参考数据定义和数据标准，从而保证数据在各个环节的拉通。

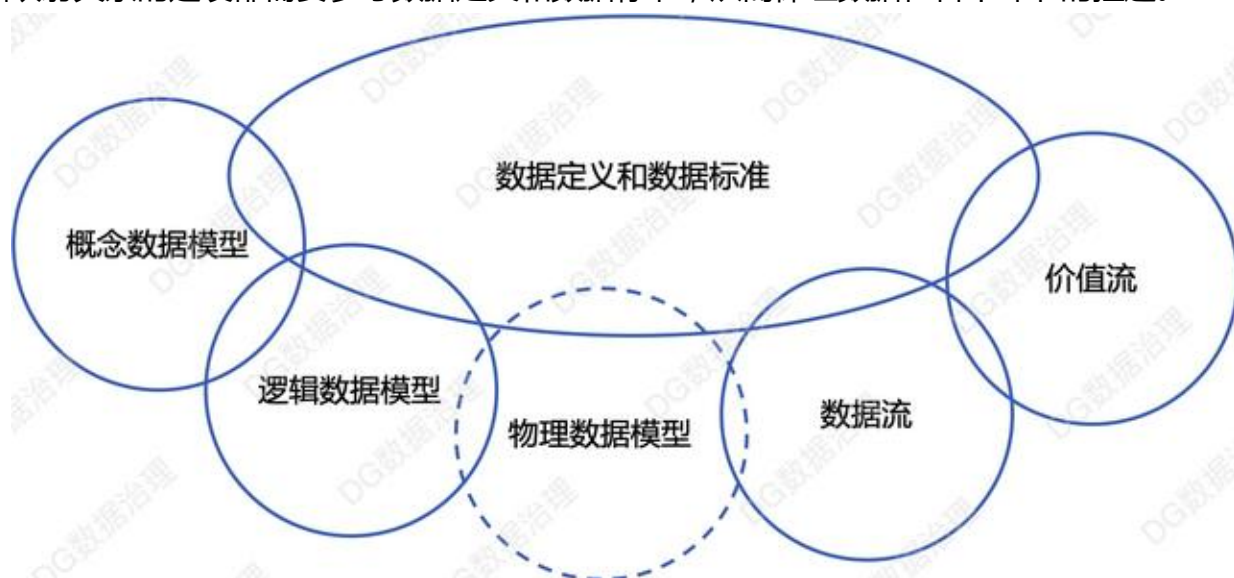


图3 数据架构的关键组件

## 4、数据架构的发展趋势？

现在数据中台建设已经成为企业数字化转型的必选项，数据架构需要从模型为导向，走向以服务为导向。传统数据架构更重视内部信息的建模和梳理，为信息化打下了很好的基础。未来数据建模依然是企业数据架构的基础，这点毋庸置疑，但是为了使数据架构真正落地，为数据消费端服务，未来在数据建模的同时需要考虑数据服务需求，提前确定服务水平协议（SLA，Service Level Agreement），这将会成为未来企业数据架构规划的重点。

### 1、从关注底层模型，向关注数据服务转变

数据建模与数据集成依然是数据架构的重要任务，但数据集成方式应该参考服务化架构，不应该再单纯关注底层模型，以梳理现状为主，而应该从数据与业务的关联入手，在模型层之上规划需要提供给业务的数据服务，在数据模型设计时就要让业务明白数据服务与底层数据的关系。

### 2、从单纯规划内部数据，向规划内外部数据集成与交互转变

客户数据、采购数据、互联网数据、IOT数据等对数字化企业越来越重要，数据架构需要重点规划外部的数据模型，以实现企业内部数据与外部数据在模型层面的连接。自GDPR（通用数据保护条例）发布以来，由于数据安全的收紧，外部数据的整体数据流也需要详细规划。

数据架构的演进远远没有结束，伴随未来DataOps等场景的出现，数据架构这部分内容会有更多的变化和新发现，数据架构也会越来越走向最终的数据消费端。

## 5、参考

[1]DAMA International.DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition)[M].Technics Publications:BASKING RIDGE, NEW JERSEY,2017:97.

[2]Steve Hoberman.Data Modeling Made Simple: A Practical Guide for Business and IT Professionals[M].Technics Publications:BASKING RIDGE, NEW JERSEY,2009[3]天行

健-任我行.企业架构---几种架构的集成

[EB/OL].[http://www.360doc.com/content/11/0428/03/170126\\_112833730.shtml](http://www.360doc.com/content/11/0428/03/170126_112833730.shtml),2011-04-28.

[4]人月神话.应用架构和技术架构

[EB/OL].[http://blog.sina.com.cn/s/blog\\_493a84550101cfen.html](http://blog.sina.com.cn/s/blog_493a84550101cfen.html),2013-01-26.

[5]孤独剑0001.概念数据模型、逻辑数据模型、物理数据模型

[EB/OL].<https://blog.csdn.net/gdj0001/article/details/80137496>,2018-04-30.