

智慧城市视角下的数据中台建设探析

——以安徽为例

孟文涛

山东大学管理学院, 山东 济南

收稿日期: 2022年7月10日; 录用日期: 2022年8月10日; 发布日期: 2022年8月16日

摘要

为进一步建设智慧城市, 安徽省政府提出建设贯穿数据全周期的大数据平台, 通过“汇、治、用、管、看、评”, 实现对数据的统一集中管理。本文将以此为基础, 从“汇、治、用、管、看、评”的新角度重新解构数据中台, 提炼出实践视角的六维架构。从“标准化和组织协调”的角度探析数据中台建设的障碍并提出解决方案, 为数据中台建设的落地提供了更贴合实际的理论和结构指导。

关键词

智慧城市, 数据中台, 数据底座, 数字中国

Exploring the Construction of Data Center in the Perspective of Smart City

—Taking Anhui as an Example

Wentao Meng

Management College, Shandong University, Jinan Shandong

Received: Jul. 10th, 2022; accepted: Aug. 10th, 2022; published: Aug. 16th, 2022

Abstract

In order to further build a smart city, Anhui provincial government proposes to build a big data platform throughout the whole cycle of data, and realize the unified and centralized management of data through “convergence, governance, usage, management, supervision, and evaluation”. Based on this, this paper will reconstruct the data center from the new perspective of “convergence, governance, usage, management, supervision and evaluation” and propose a six-dimensional archi-

ecture from the perspective of practice. From the perspective of “standardization and organizational coordination”, this paper analyzes the obstacles in the construction of data center and puts forward the solutions, which provides more practical theoretical and structural guidance for the landing of data center construction of data center construction.

Keywords

Smart City, Data Center, Data Pedestal, Digital China

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

数据中台是智慧城市建设的中枢引擎，是贯通城市不同业务系统、提高管理效率和数据利用效率的关键桥梁。在信息技术不断发展的今天，技术障碍已经不再是数据中台建设的最大困难，事实上，数据中台从设想到落地，势必会面临诸多现实的管理协调问题。现有的数据中台建设架构的问题在于多着眼于技术过程的解构，而轻视了实践中的管理、协调、反馈等需求。本文从实践视角重新解构数据中台，从“汇、治、用”的角度重新梳理了数据中台的功能组件，从“管、看、评”的角度设计了数据中台建设的控制组件，以此解决数据中台理论架构与实际建设中的管理需求不适配的问题。

2. 研究背景

2.1. 技术背景

作为智慧城市的数据底座，数据中台具备全流程、全方位的数据支持能力，在智慧城市的建设中发挥重要作用。随着数据挖掘、采集、存储、处理分析等技术的不断进步，数据中台建设将一步步走向完善，在智慧城市的建设中发挥更大的作用。

2.2. 实践背景

“数字中国”建设的全面铺开，“智慧城市”建设试点全面落地。目前，全国已形成包括以上海、杭州、南京、合肥为主的长三角区域城市群在内的四大智慧城市集群。

2.3. 案例背景

以合肥为中心，在中国科技大学和科大讯飞等院校和企业的技术支持下，安徽积极开展智慧城市建设，取得了颇为可观的建设成果，并积累了丰富的相关经验，其成功经验对其他地区的智慧城市建设具有一定的参考意义。

3. 相关概念解析

3.1. 智慧城市

智慧城市是在现代信息与通信技术的高速发展，以及“城市病”下需要提高现代城市治理水平的基础和推动下诞生并发展起来的[1]。智慧城市将云计算和物联网等新一代信息技术运用在城市中各行各业，以实现工业化、城镇化与信息化的深度融合，使城市居民能够快捷地、智慧地管理生产和生活状态，促

进城市的经济、社会、资源、环境的协调可持续发展，缓解“城市病”以及城市发展中的各种矛盾。

3.2. 数据中台

数据中台是一套可持续的、集成化的、高效率的城市数据中枢。可以通过提供数据上云、数据治理、数据分析、数据决策、结果展现等功能，挖掘数据价值、提高数据质量、提升数据使用效率[2]，实现数据资源管理、潜在规律挖掘、业务决策优化。最终目标是为了向城市内的各类业务应用输出智能的、标准的、全量的、干净的数据[3]。

4. 数据中台的架构

城市数据中台聚焦于真正把数据“统起来”、“管起来”、“用起来”的目标。从设想、理论到实践的不同阶段，数据中台可以被解构为以下三种架构。

4.1. 数据中台的逻辑架构

从逻辑架构上来看，数据中台就是从已有数据库、分散的组织数据平台以及个人用户处采集数据，经过技术处理形成集成的数据，通过集成数据回流组织数据平台、数据服务门户的建设、数据报告的输出分别达成同各数据平台、个人用户、决策者以及其他调用者的数据共享。如图 1 所示。

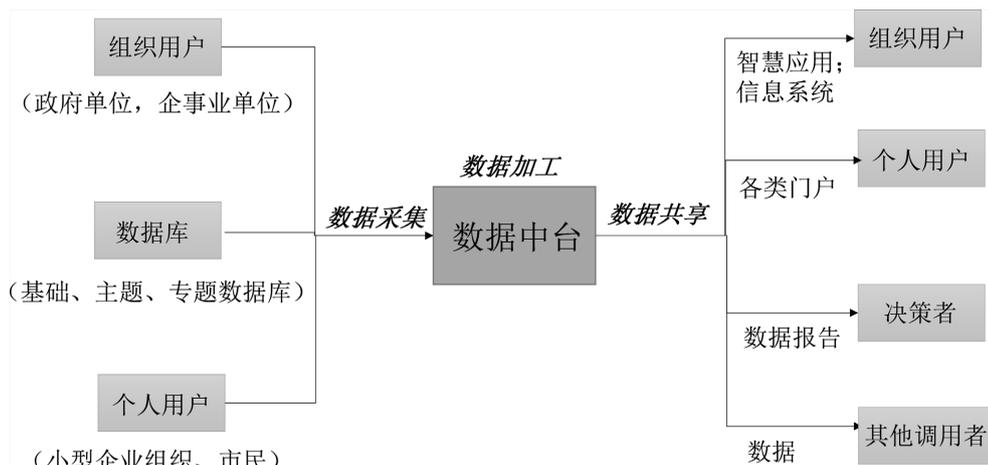


Figure 1. Logical architecture diagram of the data center
图 1. 数据中台的逻辑架构图

4.2. 技术导向的三维架构

依据数据中台的逻辑架构，结合相关技术，可以进一步解析出包括数据采集、数据加工和数据共享[4]在内的技术层面的三维架构。如表 1 (表中箭头表示数据流向)所列。这是现有研究中的主流架构，也是现有的大多数数据中台都会参考沿用的架构。

首先，来自政府部门、企事业单位、个人用户等数据源的数据被采集至数据中台。然后在数据中台中对数据进行加工处理，需依次完成以下步骤：1) 利用数据存储、计算与搜索引擎、机器学习和数据分级分类等技术对所采集的数据进行汇集；2) 通过元数据管理、数据建模、数据血缘管理等技术对所汇聚的数据集进行数据治理，提高数据质量；3) 将治理后的数据分类融合，归入不同的主题库、专题库、系统数据库等；4) 利用大数据分析、知识图谱等技术对分类后的数据进行分析，形成可供决策支撑的信息流或报表[5]。最后，将数据从数据中台导出至数据共享平台、数据开放平台和门户网站等外部应用，实

现数据的共享。

Table 1. Technology-oriented 3D architecture

表 1. 技术导向的三维架构

外层应用	数据共享平台、数据开放平台、门户网站、APP 等
数据中台	↑ 数据共享(共享至外层应用)
数据分析	大数据分析、知识图谱、分析报表、决策支撑
数据融合	主(专)题库、政务信息库、系统数据库、日志数据集、数据仓库
数据治理	基础数据管理、数据建模、数据资产管理、元数据管理、数据血缘管理
数据汇聚	数据存储、计算引擎、数据分级分类、搜索引擎、机器学习
	↑ 数据采集(采集自外层数据)
外层数据	政府单位、企事业单位、个人填报、其他数据库等

4.3. 实践视角的六维架构

数据中台的建设不仅仅是一个技术问题，更是一个管理实践问题，不能仅仅单纯从技术层面进行解析，还需顾及管理者、使用者等主要参与者，结合实际的组织和管理问题。“汇、治、用、管、看、评” [6]作为数据中台建设的标准，除了技术指标，还提供了用户、管理者的角度，可以用于解析实践领域数据中台的架构，作为技术导向的三维架构在实践领域的延伸和细化。其具体维度和解析如表 2 所述。

Table 2. Dimensions and resolution of six-dimensional architectures

表 2. 六维架构的维度和解析

维度	解析
汇	包含底层的数据汇聚以及大数据湖的建设等
治	按照数据治理规范对大数据湖的数据进行治理、整合以及基础库、主题库的建设
用	以平台用户为出发点，提高总门户、开放门户、数据应用等数据使用能力
管	从管理者角度出发，进行数据管理、标准规范建设、安全体系建设等
看	对数据进行全过程监控，让数据“看得见、看得清、看得懂”。主要包括智能数据驾驶舱、数据展示调度的能力
评	从数据的不同维度对数据的各提供部门单位考核评估。通过数据评价考核机制倒逼政府各部门协同推进大数据发展

在“汇、治、用、管、看、评”六维架构(如图 2)中，涵盖了“数据采集、数据加工、数据共享”的三维架构。即“汇、治、用”。先利用数据集成、数据交换和存储计算平台汇集数据形成大数据湖；大数据湖中的数据经过加工(数据治理、数据调度、数据开发、标签管理和实时计算)后，融合形成各类基础库和主题库；再通过市级子门户、共享数据平台、数据沙箱、工序协同平台、市级数据开放平台实现数据共享。此外，“看”、“管”、“评”三个维度同样从“数据采集、数据加工、数据共享”的角度监控、管理和评价全过程的数据流动。以此实现技术和实践的融合，协调实践中可能遇到的水平和垂直方向的冲突。

事实上，“汇、治、用、管、看、评”六维架构可以更形象的表述成一辆汽车上的六个重要组件(如

图3)。“汇”、“治”、“用”是动力组件，“汇”是油箱，提供驱动数据中台的原动力(数据集)，“治”是引擎，将油箱中的能源转化成可以直接使用的能量(加工处理后的数据)，“用”是轮胎，借用前两步产生的能量，发挥数据中台的作用(实现数据共享)。“管”、“看”、“评”是控制组件，“管”是主驾驶，是直接控制手段，管控数据中台的运行；“看”是副驾驶，是间接控制手段，监控数据中台运行的全流程；“评”是车牌号，是外部控制手段，通过标准化的要求和考核，指导数据中台的建设和发展。

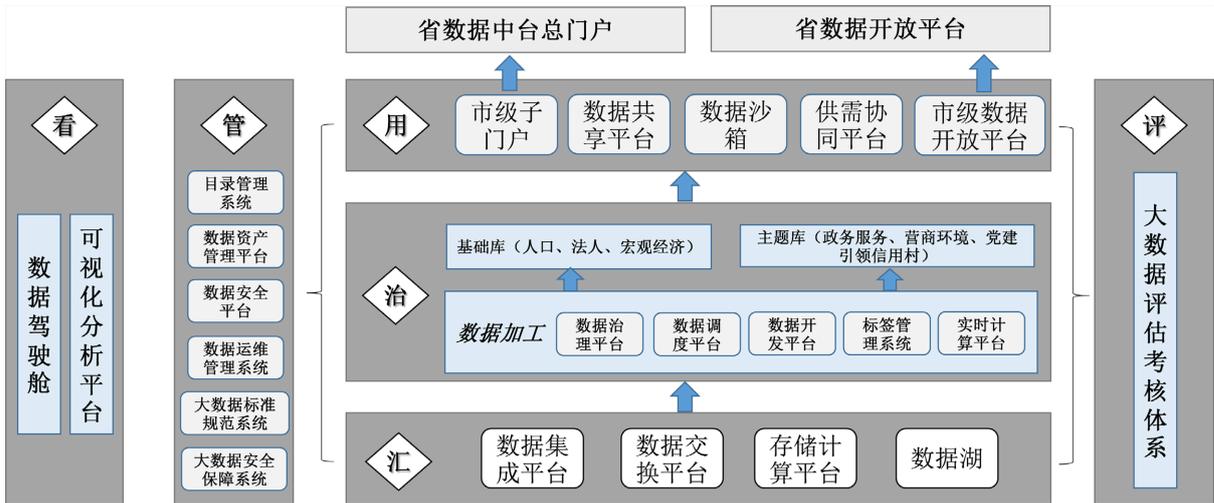


Figure 2. A six-dimensional architecture diagram from a practical perspective
图 2. 实践视角的六维架构图

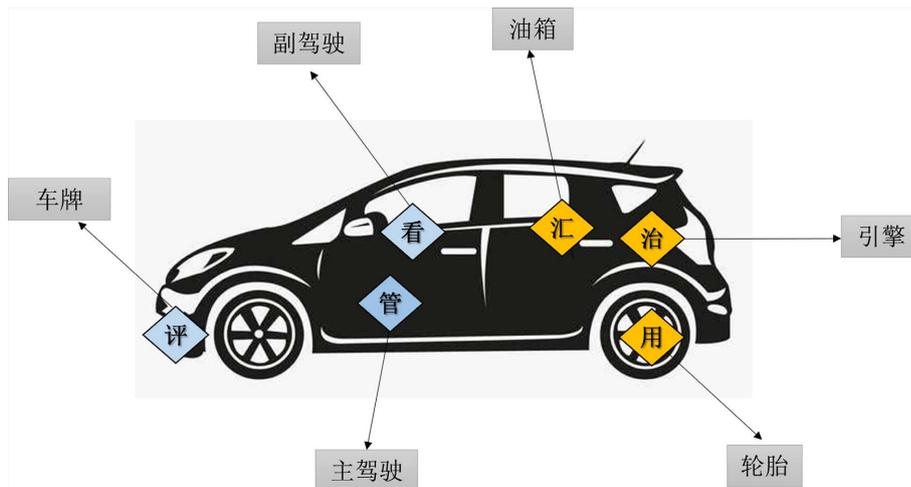


Figure 3. Image display of six-dimensional architecture
图 3. 六维架构的形象展示图

4.4. 六维架构各维度关键技术解析

此部分将对六维架构中各个维度关键组件的功能和技术进行解析。

4.4.1. 汇

1) 存储与计算平台：基于 Hadoop 框架的私有化平台，具备可靠、安全、容错、易用的集群管理能力，适配国产化软硬件，标准化性能指标。为数据中台以及业务系统提供计算与存储资源，包括离线计

算、流式计算、内存计算、分布式文件系统、分布式数据库等。

2) 数据集成平台: 消除数据孤岛的同步平台, 为各业务方提供数据接入的高效通道。可以实现异构数据源互通集成, 包括各种关系数据库、大数据组件, 实时数据, 文件数据及接口数据等。支持集成过程中的映射、转换、标准化等预处理, 同时提供有效监控服务, 实时分析任务运行调度的情况。

3) 数据交换平台: 由一个可视化的数据查看首页以及前置机管理、数据源管理、任务管理、监控管理、系统管理组成, 可实现部门间以及省、市、县三级平台间数据交换, 致力于消除部门间、地域间、层级间数据共享困难、信息不一致等问题。

4) 数据湖: 基于 Hadoop 框架, 为了应对不同业务需求的特点, 采用传统数据仓库架构 + MPP 数据库 + Hadoop 集群的统一存储平台。

4.4.2. 治

1) 数据治理平台: 提供错误数据检查组件(空数据、缺失数据等)和数据清洗转换处理组件(空数据补充、格式统一、数据编码统一等), 对问题数据进行处理, 生成正确的、标准一致的可用数据, 为上层数据分析应用提供保障。

2) 数据调度平台: 提供数据接入、数据检查、数据治理、数据处理等任务的编排调度管理能力, 并支持多任务并发运行, 监控调度运行中的所有运行日志信息, 提升系统开发和维护的质量, 减少系统建设、实施及运维等各方面成本。

3) 数据开发平台: 内置多种机器学习算法, 设定多个预处理节点, 并开放自定义算法入口供用户数据分析使用。

4) 标签管理平台: 按照数据实际的标签体系, 对数据治理工程中的关键数据进行业务抽象, 将数据以指标或标签定义出来, 提供给治理平台、开发平台以及其他业务系统使用。

5) 实时计算平台: 基于事件驱动的轻量级流计算引擎, 提供数据实时采集、实时计算、实时集成的全链路流计算。内置丰富的流计算组件, 提供可视化的作业编排; 支持通过 SQL 的方式实现流式数据开发, 大幅降低了流式数据分析门槛。

4.4.3. 用

1) 数据沙箱: 数据沙箱对数据进行订阅和脱敏管理并提供沙箱环境, 支持数据载入、镜像管理、文件上传、结果数据导出, 在通用机器学习框架下进行数据分析挖掘, 实现数据“可用不可见”, 支撑数据安全流通和分析挖掘数据价值。

2) 市级子门户、数据共享平台、供需协同平台、市级数据开放平台: 针对不同用户和数据使用者, 是数据共享的直接门户。

4.4.4. 管

1) 目录管理系统: 盘点梳理业务数据, 编制发布业务目录, 规划指导数据的接入、处理、管理、共享。

2) 数据资产管理平台: 将数据标准(标准字典、数据源字典、数据元)进行统一管理, 为数据治理提供标准依据; 将元数据信息(技术元数据、操作元数据)通过可视化方式管理, 利用管理的元数据进行数据目录创建; 同时提供主数据对象以及主数据维护界面, 并支持以服务的方式对外提供。

3) 数据运维管理平台: 通过查看总体运行监控、资源目录监控、数据资源监控、数据共享监控、数据采集监控、调度任务监控、数据服务监控、资源更新监控等, 全方位、可视化监控数据中台整体运行情况。依据展示系统中异常、失败、预警等可视化展示快速发现问题原因及位置。

4.4.5. 看

1) 数据驾驶舱:通过数据建模分析、可视化等技术,挖掘各单位业务领域核心方向和绩效考评指标,围绕城市运行的关键领域,以高度可视化、图形化的方式实时展示数据运行体征数据。不仅对数据全生命周期进行监测、预警、分析、评价、指挥,实现全周期管理,同时对重点主题领域的运行状况进行直观展现,为政府的政务决策提供数据依据。

2) 可视化分析平台:基于多种数据源(mysql/gbase/oracle/dameng/MQ/地理数据源)对数据进行接入及处理,内置多种素材、模板、内置组件(趋势图、占比图、GIS 地图等),辅助用户进行数据编排、样式配置及成果展示与分享。

4.4.6. 评

1) 大数据评估考核系统:通过多维度(数据共享、数据规范、数据质量、数据应用等)搭建考核评估指标,制定考核方案,依托模型计算从而获取评估结果,并进行多维度可视化对比分析,以评估评价考核对象各项常态化工作情况。

4.5. 当前主流架构和六维架构的对比

现有的研究成果多是基于逻辑架构具体解析出的“采集-处理-共享”三维架构,区别主要在于具体技术和功能组件的不同。也有部分学者将管理和评价纳入主流架构,但均没有涉及细节,也没有设计具体的功能组件和协调机制,在实践方面缺乏指导意义。将主流的三维架构和实践视角的六维架构进行多层次的对比,参见表 3,可以在一定程度上体现出六维架构在实践方面的指导意义。即:实践视角的六维架构更多地考虑到了用户、管理者等相关群体,立足于实际,注重于实际中组织、管理等问题的协调和处理,能够在数据中台建设落地过程中发挥方法论的指导作用。

Table 3. Three-dimensional versus six-dimensional architectures

表 3. 三维架构和六维架构的对比

	三维架构	六维架构
视角	开发者	用户、管理者
导向	技术	实践
关注点	数据的收集、加工和共享	实际问题的协调和处理
潜在假设	技术问题是建设数据中台建设的核心	组织管理与协调等实际问题同样举足轻重

5. 数据中台的实践障碍

5.1. 数据标准化

数据中台涉及众多数据、众多业务系统,不同系统间的不同数据格式,会影响数据质量和处理效率。此外,传统数据仓库使用 ETL 将业务系统的数据直接采集加工至数仓中进行数据整合,在应对不断增加的业务量时,会造成数据混乱,无法满足数据标准。

5.2. 标准制定与协调问题

数据中台建设是一项新兴的、多部门、长线程的工程,具有技术性强、前瞻性高、涉及部门广泛、可供参考的经验少和相关综合管理人才不足等特点。故很容易产生评价标准的制定不合理和组织协调困难等问题。比如,评价标准的制定没有充分考虑技术限制、由于忽视数据标准化建设的难度,缺乏针对

各部门的专业指导，导致数据标准化实施成效低等。

6. 解决措施

6.1. 分层处理数据

数据应当依照数据中台原则分层设计，自下而上可以依次分为：源数据层、中间层、应用层。由于各层次之间的数据具备低耦合和高内聚的特征，可以尽量避免因业务系统过多而造成的数据混乱[4]。

6.2. 提供坚实的组织保障

完善组织保障体系，建立跨部门的领导小组。包括综合协调组(市领导、业主单位、平台承建厂商)、标准规范组(业主单位、平台承建厂商)、业务组(公安、民政等政府业务部门、业主单位、平台承建厂商)、数据技术组(平台承建厂商、部门业务系统厂商)和平台建设组(平台承建厂商)。群力群策制定统一的、灵活的标准规范体系，包括统一的管理办法、技术标准和数据标准。

7. 结语

本文基于安徽省数据中台建设的经验，结合数据中台的逻辑架构和技术架构，有别于主流架构，从实践的视角，重新梳理了数据中台的总体架构，从“汇、治、用、管、看、评”多维度设计了数据中台的功能组件和控制组件，为数据中台建设的落地提供了更贴合实际的理论和结构指导。同时还指出数据中台建设实践中的一些问题并提出解决方案，可以为其他城市的数据中台建设提供一定的参考。

参考文献

- [1] 邢玉杰. 新时期智慧城市建设与发展的路径研究[J]. 财经界, 2022(10): 74-76.
- [2] 张海峰. 智慧城市数据底座建设探析[J]. 智能建筑与智慧城市, 2022(3): 149-151.
- [3] 付登波, 江敏, 任寅姿, 等. 数据中台—让数据用起来[M]. 北京: 机械工业出版社, 2020.
- [4] 许苗峰, 薛慧. 智慧城市数据中台设计与应用[J]. 邮电设计技术, 2020(2): 84-86.
- [5] 王智, 尹长林, 许文强. 智慧城市背景下数据中台的研究与设计[J]. 网络安全和信息化, 2021(7): 29-35.
- [6] 安徽省人民政府. 安徽省人民政府关于印发安徽省“数字政府”建设规划(2020-2025 年)的通知[J]. 安徽省人民政府公报, 2020(21): 3-30.