

# 园区大数据服务平台的研究与设计

阳梦华

(湖南省邮电规划设计院有限公司,湖南长沙 410000)

**【摘要】**随着大数据技术的日益成熟,数据成为了新时代重要的生产要素。分析园区数据平台的主要建设需求,基于“以共享为基础、以需求为导向”的原则,搭建园区大数据服务平台架构,对园区内业务数据资源进行全面盘点,形成园区数据资产,通过大数据服务平台的建模和挖掘分析,帮助园区进行企业洞察分析、产业链分析等,为园区的长远发展和健康发展提供决策依据。

**【关键词】**大数据;数据要素;产业链

**【doi:10.3969/j.issn.2095-7661.2023.01.010】**

**【中图分类号】**TP311.13

**【文献标识码】**A

**【文章编号】**2095-7661(2023)01-0040-03

## Research and Design of Big Data Service Platform in the Area

YANG Meng-hua

(Hunan Planning & Designing Institute of Post & Telecommunication Co., Ltd., Changsha, Hunan, China 410000)

**Abstract:** With the growing maturity of big data technology, data has become an important factor of production in the new era. This paper analyzes the main construction needs of the area's data platform, builds the area's big data service platform architecture based on the principle of "sharing-based and demand-oriented", conducts a comprehensive inventory of the business data resources in the area, and forms the area's data assets. Through the modeling and mining analysis of the big data service platform, the area carries out enterprise insight analysis, industry chain analysis etc., which provides decision-making basis for the long-term and healthy development of the area.

**Keywords:** big data; data elements; industrial chain

随着数字经济的发展,数据成为战略性资源,成为新生产要素<sup>[1]</sup>,各个区域和业务领域对数据的重视日益加强。大数据成为新型生产资料,正在给人类的生产、生活带来翻天覆地的变化,正在改变社会生产方式和人们的生活方式<sup>[2]</sup>。数据资源作为一种跨行业、跨部门、跨类型的资产,具有重要的公共价值和经济价值。如何评估数据的价值和有效性,是关系到数据治理方向和治理质量的重要环节<sup>[3]</sup>,因此搭建大数据服务平台和数据治理体系,有助于有效挖掘并充分发挥数据资产的价值。

园区大数据服务平台应注重园区信息战略定位,站在园区数据运营的角度,通过大数据服务的建设,有效推进数据资源的整合共享,避免由于各

单位独自建设导致规范标准不统一、建设内容有重叠、投资浪费等现象<sup>[4]</sup>。而大数据作为信息化的数据资源和技术手段,对园区经济产业发展具有巨大价值<sup>[5]</sup>,数据是基础性、战略性的资源,提高数据资源的利用价值已经成为各园区未来健康持续发展的前提和保障。

### 1 园区数据平台建设需求

#### 1.1 明确数据来源,加强数据采集

园区数据平台的建设需要明确园区内部数据资源的分布情况,大规模数据资源的接入,应根据大数据分析需求对园区内部数据及外部企业相关数据源头进行明确,通过数据采集系统不断丰富数据资源。

**【收稿日期】** 2023-01-10

**【作者简介】** 阳梦华(1988—),女,湖南衡阳人,通信工程师,硕士,研究方向:大数据技术。

## 1.2 打通数据壁垒,实现数据共享交换

不少园区现已打通数据资源对接,但大多是以点对点的方式进行数据对接,且对接数据资源不全,还未完全实现数据的互联互通,数据壁垒还是存在,因此需要建立共享共用的数据共享交换系统,通过前期梳理的园区内的数据资源目录,为园区内部数据资源共享提供数据通道。

## 1.3 加强数据管理,建立数据治理体系

园区内数据标准、数据质量等存在不规范,需要对归集的数据资源进行标准化管理,建立元数据管理、质量管理、数据血缘管理及数据安全管理体系,不断规范和提升数据质量,对外输出优质数据,为数据分析和数据应用提供保障。

## 1.4 加快数据处理,搭建大数据服务平台

随着园区企业、业务及产业链的增多,园区内数据规模也在增大,为加快数据处理,满足园区管理决策需求,须搭建基于Hadoop体系的大数据服务平台,利用数据采集、清洗、加工及存储、计算、分析等工具,对园区各类数据进行建模和分析。

## 2 园区大数据服务平台建设内容

园区大数据服务平台基于“以共享为基础、以需求为导向”的原则,通过数据资源的采集、共享共用,加强数据资源管理和服务能力,形成大数据资源池,建立统一的数据治理体系,依靠大数据基础框架,提升数据资产价值,为园区内各单位提供大数据分析服务,平台架构如图1所示。

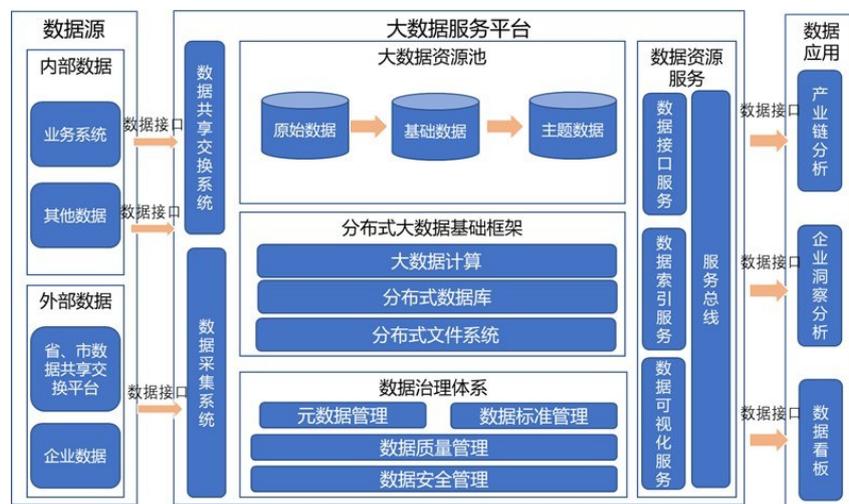


图1 园区大数据服务平台架构图

### 2.1 数据采集系统

数据采集系统主要包括离线数据采集、实时数据采集及互联网数据采集等,通过数据接口方式接入多种数据类型,比如:结构化数据、半结构化数据和非结构化数据等。同时支持全量数据、增量数据等接入策略,在不同的业务场景下,可选择不同的接入策略,提高数据接入效率。

### 2.2 数据共享交换系统

数据共享交换系统主要是对园区各单位的业务系统进行打通,通过数据申请、订阅等方式,实现园区内各单位数据资源的共享和交换,可有效节约人力成本。

数据共享交换系统的实现是通过各单位将共享交换的数据推送到前置库中,并将共享的数据录入到数据资源目录中,需要使用数据的单位通过数据目录进行订阅和申请,待共享数据的单位审核同意后,使用数据的单位则可通过数据接口接入相应的数据。

### 2.3 数据治理体系

数据治理体系的搭建是基于数据资源的统一管理,不仅能提高园区的数据服务质量和效率,保障大数据服务平台数据资源池的权威性,也能为园区数据应用提供高质量的数据,实现数据内部充分共享与对外开放合作,更好地为园区宏观调控及管理决策提供服务。治理体系主要包括:

1)元数据管理主要包括技术元数据、业务元数据和管理元数据,主要包括对元数据的注册、编目、存储和发布。同时建立元数据模型,支持元数据的导入和导出,并可对模型进行检查和比对,有利于对元数据模型的维护和管理。

2)数据标准管理是通过建立一套数据标准规范,制定企业的各项数据标准内容,提供录入、管理以及维护标准的功能,主要建立公共数据元标准,确定数据元数值范围、数据类型、长度等,统一数据源口径,建立数据标准管理办法,防止数据资源的混乱使用,保证数据质量准确性和规范性,并提供数据标准查询入口,具备对数据标准全生命周期的管理能力。

3)数据质量管理是基于数据标准规范和质量规则基础上,依托数据质量规则引擎,对数据质量管理规则进行配置,通过质量监控、数据稽核、问题反馈、数据审核、数据质量报告等环节来达到数据质量的管控。在数据质量方面,必须要保证数据的完整性、一致性、准确性和及时性等。

4)数据血缘关系管理基于元数据管理,在数据产生、加工融合、流转流通到最终消亡等处理过程中,建立数据资源的血缘关系。同时能够对数据资源库的类别、数据项级别进行数据血缘追溯。在数据治理的过程中,血缘监控需要根据数据项的治理过程来达到血缘监控要求。并且通过不同的数据类型,需对数据资源库进行相关流向分析和数据溯源操作。

5)数据安全是数据治理体系的核心,数据的安全是大数据服务平台搭建的基础,主要包括数据存储、传输、使用、访问等安全管理。

## 2.4 大数据基础框架

大数据基础框架主要是采用分布式大数据平台,支持结构化、非结构化、半结构化数据的存储、加工、处理以及计算,提供平台级资源和数据管理能力,能够很好地整合园区内各单位数据,对接相关数据及外部企业相关数据,提供数据清洗和加工工具,按照园区法人、人口、产业、经济、资产等主题进行数据建模。利用数据的分析挖掘与可视化技术、集成工具和算法,实现一站式数据分析。

## 2.5 数据资源服务

数据资源服务是园区大数据服务平台提供数据访问和管理能力的服务,统一面向业务人员提供数据资源管理、维护和发布,使园区各用户及时感知到数据资源,快速搜索到符合自己需求的数据,同时向用户提供数据间的多维关系自动化识别和可视化展示。主要包括数据资源目录服务、数据查询检索服务、数据报表服务、数据分析服务、数据标签服务、数据可视化服务、数据接口服务。

## 3 园区大数据应用场景

### 3.1 企业洞察分析

企业洞察分析通过归集企业数据资源,通过大数据分析建模技术,实现对园区内产业宏观分析、企业生长力评价、企业画像、企业体检、企业预警等功能,为相关管理部门提供全面了解区域内企业生产经营状况的技术手段,为园区更好地进行产业服务、响应国家对民营经济扶持政策提供重要支撑。

1)企业生长力评价。基于企业生长力评级体系,以国内外关于企业竞争力和经营发展能力评估等理论与方法为基础,基于园区大数据服务平台,基于指定区域内经济、社会环境、企业等海量数据,利用企业经营数据分析和多维度画像,整合企业生长力的各类指标,搭建企业生长力指标体系,描绘企业动态的生长力指数。力求全面、客观、准确地反映一个企业成长过程的变化特点,形成一套比较完整的指标体系和评估方法。

2)企业画像。提供入驻企业数据分析结果的可视化展示,精细化呈现企业经营状况。引入新经济指标体系,获取所属区域的企业经营状况、查看各类分析报告,真正做到数据随身、决策随行,高效便捷地实现移动决策。主要包括企业评级分布分析、企业生长力指数分析、企业分布行业分析、企业定位、企业详情、企业生长力报告。

## 3.2 产业链分析

产业链分析是基于大数据服务平台归集整合园区内相关企业的数据库资源,按照园区现有产业链的维度进行分类,关注每条产业链相关企业的上下游情况,以及各企业的区域分布等,旨在更好地掌握和了解各产业链的建设和发展。通过产业链的统计分析,对园区内各产业链的上中下游企业进行分类,对各产业链上、中、下游企业数量、企业运行状态、存在问题及是否健康发展进行统计和分类,更好地辅助园区领导层进行产业链的布局和调整,更加持续健康地推进产业链的发展。

## 4 总结

基于园区的大数据服务平台的建设是对园区数据资源进行归集整合的基础,通过大数据平台的建模和挖掘分析,帮助园区进行企业洞察分析、产业链分析以及经济指标等分析,为园区的健康长远发展提供决策依据。

## 【参考文献】

- [1]李海舰,赵丽.数据成为生产要素:特征、机制与价值形态演进[J].上海经济研究,2021(8):48-59.
- [2]张建云.大数据技术体系与当代生产力革命[J].马克思主义研究,2021(4):58-68,164.
- [3]朱泽,段尧清,何丹.面向政府数据治理的数据资产价值系统仿真评估[J].图书馆论坛,2021(6):100-105.
- [4]刘春林,徐文辉.市级政务大数据平台建设思路[J].信息技术与信息化,2020(3):173-176.
- [5]刘艳,张肖会,吴丽.大数据与产业融合发展[J].天津科技,2021(12):4-6.