

大数据资源应用

Massive datasets resource application

(征求意见稿)

(本草案完成时间: 2021-8-21)

在提交反馈意见时, 请将您知道的相关专利连同支持性文件一并附上。

XXXX - XX - XX 发布

XXXX - XX - XX 实施

目 次

前言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 综述	1
4.1 数据资源	1
4.2 承载形式	2
4.3 存在形态	2
4.4 应用模式	2
5 基本架构	3
5.1 总体架构	3
5.2 数据资源服务节点	3
6 规划设计	5
6.1 要求	5
6.2 顶层规划	5
6.3 模块化设计	6
7 评估	7

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由辽宁省工业和信息化厅提出并归口。

本文件起草单位：大连软件行业协会、大连交通大学、大连软信咨询服务有限公司、辽宁省信息中心、沈阳市信息中心、辽宁省社会保险事业服务中心、信华信（大连）软件服务股份有限公司、大连奥远电子股份有限公司、大连理工现代工程检测有限公司、大连市计算机学会。

本文件主要起草人：郎庆斌、尹宏、于青、刘强、唐文伟、明晓浙、邓连峰、于昕、丁宗安、刘宏、才昊、胡剑锋、秦健、宋悦、王鑫。

本文件发布实施后，任何单位和个人如有问题和意见建议，均可以通过来电和来函等方式进行反馈，我们将及时答复并认真处理，根据实际情况依法进行评估及复审。

本文件归口单位通讯地址：沈阳市北陵大街45-2号，联系电话：024-86913384

本文件起草单位通讯地址：大连市高新园区火炬路32号创业大厦A座5层，联系电话：0411-88255657

大数据资源应用

1 范围

本文件定义了大数据资源基本架构、大数据资源规划设计的基本规则，并描述了大数据的基本样态。本文件适用于大数据资源的相关应用。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

DB21/T 1799.2 信息技术 信息服务管理规范 第2部分：IT系统集成

3 术语和定义

下列术语和定义适用于本文件。

3.1

大数据 massive datasets

海量、复杂的结构化、半结构化、非结构化、类型繁杂的规模化数据集合。

3.2

泛在网络 Ubiquitous Network

普遍的、无所不在的多种型式信息网络体系。

3.3

云 cloud

本标准所指云是基于泛在网络实现的可用、便捷、按需应用的服务模式。

注1：云计算应是基于云端展开的各种应用、服务。

3.4

云端 clouds

本标准所指云端是泛在网络形成的网络空间，即云中。

3.5

云端服务 clouds services

基于云模式提供的各种服务。

4 综述

4.1 数据资源

数据是一种资源，亦是资源的表现形式。海量、复杂的各类数据，依据一定规则相互关联构成的数据集形成规模化数据资源（大数据资源）。以数据资源为基础，研究、开发、处理、使用等的相关技术、应用、管理、服务，形成与IT相关联的跨学科、跨领域的复合应用。

跨学科、跨领域的的数据资源复合应用，需要建立研究、开发、处理、使用等的科学化、规范化、顶层化、协同化秩序，规范相关技术、应用、管理、服务的约束规则。

4.2 承载形式

由于不同领域、行业以及个人应用（业务）系统的多样性，所产生的大数据资源种类繁多，且承载这些数据资源的形式也呈现出多样化的特点，主要应包括：

- a) 媒介：
 - 1) 磁介质：计算机硬盘、数据存储设备（如磁盘阵列等）、移动存储设备（如移动硬盘、U盘、磁带等）、手持移动设备中用于数据存储的组件（如智能手机、个人数码助理中的内置存储芯片或可拆卸存储卡等）；
 - 2) 光介质：光盘、光存储设备等；
 - 3) 芯片介质：芯片卡（如银行卡、护照等）；
 - 4) 纸介质：纸质文档；
 - 5) 电子媒介：广播、电视、电影等；
 - 6) 多媒体媒介：录音、录像等；
- b) 网络：博客、微博、微信、论坛、网盘、云盘、邮件、即时通讯、网站、网络视频等。

4.3 存在形态

大数据资源应用形式的多样性，形成不同的数据资源存在形态，主要应包括：

- a) 离散的：分散存储或保存的数据，无关联或可建立关联关系；
- b) 网络的：基于网络应用形成的集中存储的数据，存在关联关系；
- c) 其它形态等。

4.4 应用模式

4.4.1 本地模式

非自动处理、非网络化应用等，如纸介质承载数据的应用，单机应用等。

4.4.2 云模式

4.4.2.1 云

云应是全分布式、泛在的网络架构形式，其特征主要表现为：

- a) 云应是基于泛在网络形成的泛化的网络应用服务架构，基于有线、无线、光纤等多种网络型式；
- b) 云架构中应无主节点，无明确的主、从节点；
- c) 云架构可随着泛在网络延伸，无边界，可扩展；
- d) 云应是开放的，可共享，灵活方便；
- e) 大数据资源应用服务通过云端服务平台实现等。

4.4.2.2 模式

数据资源或散布、或集中于某些或某一个节点，数据资源在云端流动，可随时、按需使用。

4.4.2.3 超规模数据中心

当大数据资源完全集中于一点，形成超规模数据中心（云数据中心），通过云端服务平台提供数据资源服务。

5 基本架构

5.1 总体架构

5.1.1 描述

基于泛在网络的云模式总体架构，应是：

- a) 离散的多数据资源服务节点全分布无边界部署（广义）；
- b) 单节点数据资源服务规划部署（超规模数据中心）；
- c) 云端提供数据资源的无障碍通道等。

5.1.2 示意图

基于泛在网络的云模式总体架构示意图，参见图1、图2。



图1 云模式总体架构示意图 1

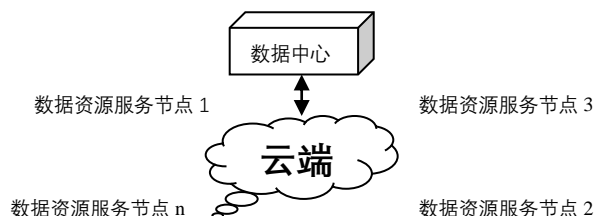


图2 云模式总体架构示意图 2

5.2 数据资源服务节点

5.2.1 描述

数据资源服务节点，应是：

- a) 提供数据资源访问、应用、调度、管理等的服务平台；
- b) 提供数据资源管理的相应存储技术（如数据资源池、数据资源仓库等）；
- c) 建立基于云端的节点互联，畅通信息通道，实现数据资源的高度共享、数据资源应用的高度协同等。

5.2.2 资源组织

基于数据资源服务节点的资源组织，主要应包括：

- a) 数据资源获取：各种形式数据资源的获取，包括实时数据采集、在线和离线数据收集、网络数据收集等；

- b) 数据资源管理：建立各类非自动化处理、自动化处理，以及结构化、半结构化、非结构化数据资源的逻辑统一的管理模式；
- c) 数据资源服务：提供网络化、非网络化数据资源相关的访问、处理、使用服务；
- d) 管理机制：建立相对完善的数据资源服务平台管理策略、管理机制等。

5.2.3 分层

数据资源服务节点可分层：

- a) 基础层：支撑数据资源服务的基础设施，或承载数据资源媒介的物理设施；
- b) 平台层：提供数据资源整合、共享、处理、部署、调用、管理，及应用协同等管理服务；
- c) 资源层：构建数据资源池，提供元数据管理服务；
- d) 采集层：数据资源采集；
- e) 交互层：数据资源访问等。

5.2.4 模型

数据资源服务节点模型，参见图3。

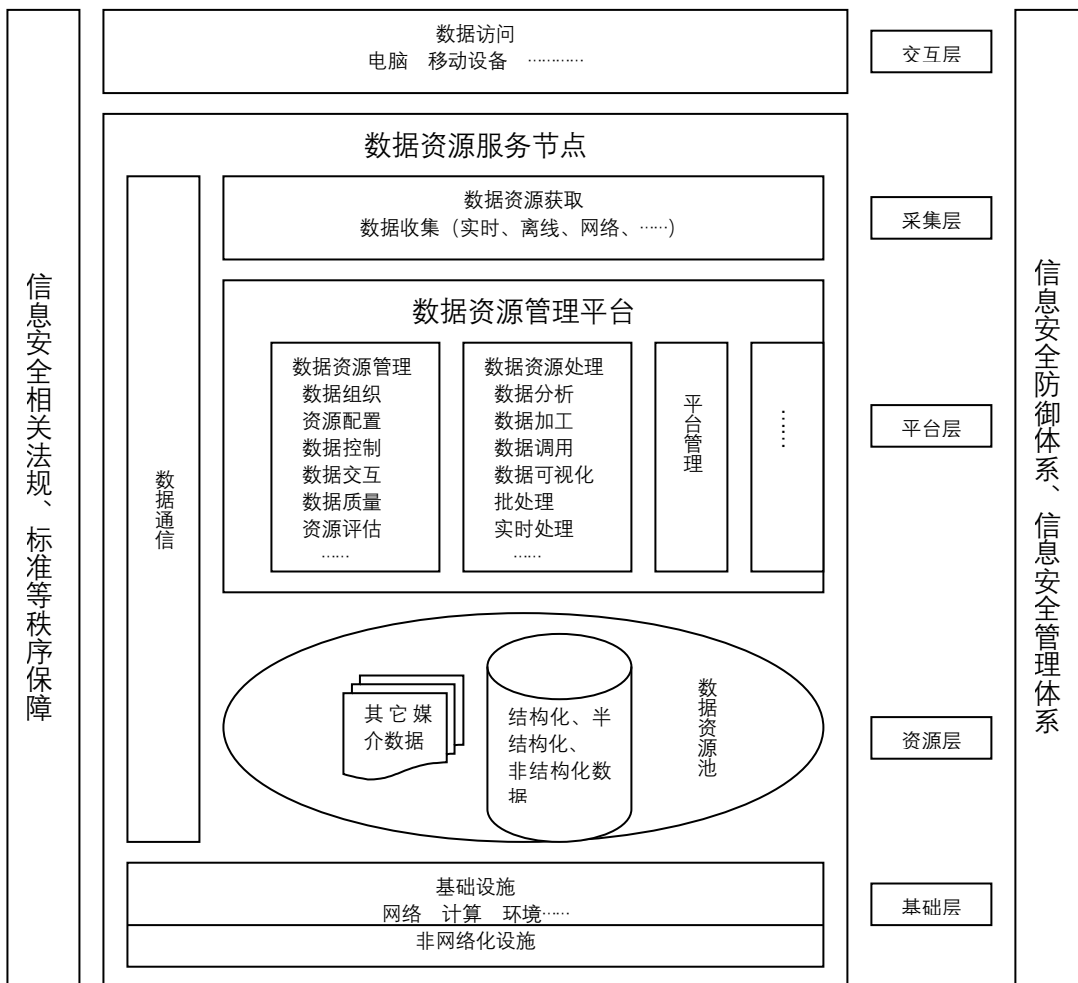


图3 数据资源服务节点模型

5.2.5 风险管理

数据资源服务节点应遵循相关信息安全法规、标准，建立风险管理和风险应对机制，实时评估安全风险，保障数据资源的安全、可靠、可信、可用。

6 规划设计

6.1 要求

大数据资源应用应科学、系统地顶层规划，并在规划框架下根据不同的应用模块化设计：

- a) 大数据相关资源应统一整合、配置、部署；
- b) 大数据资源应统一管理、调度和控制；
- c) 应根据大数据资源属性规划相应的应用；
- d) 应避免大数据的产业性开发，保证大数据的纯粹、准确、可信、可用。

6.2 顶层规划

6.2.1 原则

大数据资源应用顶层规划的原则，主要应包括：

- a) 系统性：基于大数据资源的关联性、完整性，大数据资源应用规划应科学、系统、规范；
- b) 通用性：基于大数据资源应用的总体框架，大数据资源应用应模块化、标准化、跨平台，并具有通用的数据管理功能；
- c) 前瞻性：大数据资源应用规划，应考虑社会、经济发展的战略目标和发展趋势；大数据资源的演化路径；密集技术的综合发展和成熟性、先进性、可用性等；
- d) 经济性：大数据资源应用规划，应综合利用、优化配置相关资源，平衡相关生态，以获取较高的经济、社会价值。

6.2.2 要素

大数据资源应用顶层规划、设计的主要要素应包括：

- a) 知识：大数据资源应用顶层规划、设计涉及 IT、领域或行业、大数据属性/特征、大数据相关环境/基础等的相关知识和复合应用；
- b) 资源：大数据资源应用相关资源的识别、整合、部署和优化配置；
- c) 技术：大数据资源应用相关的技术储备、应用开发、关联领域复合技术应用等；
- d) 安全：大数据资源和大数据资源应用的深层安全保障体系；
- e) 人员：大数据资源应用的能力要求（专业、技术、业务等）、人才储备等；
- f) 标准：大数据资源应用标准化建设，包括技术、管理、应用、服务等标准体系。

6.2.3 目标

大数据资源应用规划，主要应明确2个目标：

- a) 战略目标：应根据经济、社会（或组织）的发展战略，确定大数据应用定位，前瞻性规划；
- b) 发展目标：应根据经济、社会（或组织）中长期发展目标和社会、经济发展及技术变革、大数据资源演化等，明确大数据资源应用的中长期发展目标。

6.2.4 数据资源

编制大数据资源应用规划，应梳理、分析、研究大数据资源：

- a) 大数据资源属性、行业/领域特征；

- b) 大数据资源形态、分类;
- c) 大数据资源分布形态;
- d) 大数据资源结构、构成;
- e) 大数据资源可提取、可处理模式;
- f) 大数据资源的可用性;
- g) 其它关联要素等。

6.2.5 规划

大数据资源应用规划、设计，主要应考虑：

- a) 大数据资源应用的需求属性、基础、环境、条件等;
- b) 大数据资源的属性/特征、部署、关联关系等;
- c) 大数据资源应用构成要素的关联关系、逻辑关系;
- d) 大数据资源应用的协同、关联;
- e) 大数据资源应用的统一资源框架和资源调度、优化、管理平台;
- f) 大数据资源应用的平台整合、控制、管理等;
- g) 大数据资源应用管理平台的边界融合、应用协同;
- h) 大数据资源、大数据资源应用的安全属性/特征、关联因素等。

6.3 模块化设计

6.3.1 描述

应在顶层规划框架下，根据行业、领域、组织形态等模块化设计大数据资源应用：

- a) 大数据资源是依据一定规则相互关联构成的数据集;
- b) 大数据资源之间具有相关性;
- c) 大数据资源应用应考虑大数据资源的相关性;
- d) 大数据资源应用设计应考虑应用之间的相关、协同和整合;
- e) 大数据资源、大数据资源应用安全是系统、全局性的等。

6.3.2 需求

应根据大数据资源的属性特征、大数据资源应用的特征，识别、分析、分解实际、潜在需求：

- a) 需求具有行业、领域或组织形态等的不同特征;
- b) 应根据不同的行业、领域或组织形态特征，识别、分析需求、需求相关性、需求演化等;
- c) 需求管理可参考 DB21/T 1799.2 等。

6.3.3 设计

大数据资源应用顶层规划框架下，模块化设计应考虑：

- a) 社会分工、产业结构等的不同划分应用类别;
- b) 类同应用类别间的大数据资源共享模式;
- c) 不同应用类别间的关联关系;
- d) 不同应用类别间资源整合和资源共享模式;
- e) 统一资源框架和管理平台下各类应用的整合、协同;
- f) 信息安全防御体系和信息安全管理体系的规范性、系统性等。

7 评估

大数据资源应用规划和设计应适时评估，并根据评估结果改进、调整、完善，以适应发展、变化的实际需求：

- a) 顶层规划的健壮性（Robust）、可信性、可用性；
 - b) 模块化设计的适用性、一致性；
 - c) 大数据资源的识别、梳理、分类；
 - d) 大数据资源应用要素的充分性、适配性；
 - e) 大数据资源、大数据资源应用的安全风险等。
-