

《农业农村大数据存储系统功能及性能要求》征求意见稿

编制说明

一、工作简况

（一）任务来源

本标准是 2022 年经农业农村部农产品质量安全监管司批准立项，归口农业农村部大数据发展中心的农业行业标准制定任务。项目下达文件：《关于下达 2022 年农业国家和行业标准制修订项目计划的通知》；文件号：农质标函（2022）66 号；项目编号：NYB-22299；项目名称：农业农村大数据存储系统功能及性能要求；项目性质：农业行业标准制定。

（二）制定背景

党中央、国务院高度重视数字经济发展，推动实施国家大数据战略。习近平总书记多次就数字中国建设作出重要指示，“要推动互联网、大数据、人工智能和实体经济深度融合，加快制造业、农业、服务业数字化、网络化、智能化”“运用大数据提升国家治理现代化水平”。农业农村大数据是大数据战略的重要组成部分，也是我们实施乡村振兴战略、建设农业现代化的重要支撑。农业农村部相继出台了一系列政策文件，推动农业农村信息化发展，促进区域农业农村数据资源的汇聚应用。《数字农业农村发展规划（2019-2025）》（农规发〔2019〕33 号）要求，以数据资源整合、数据共享应用为途径，开展技术研发与平台建设，加速农业农村数据互联互通，推进数据融合、挖掘与应用。全国范围内的各级政府部门、涉农管理部门纷纷加快了涉农数据资源的整合、农业农村大数据中心建设，但如何有效、快速、可靠地存取日益增长的海量异构涉农数据资源，对数据存储系统的功能与性能提出更高要求。

近年来，国家高度重视大数据存储和管理标准体系建设，先后制定了《信息技术 云数据存储和管理 第 1 部分：总则》（GB/T 31916.1-2015）、《信息技术 云数据存储和管理 第 3 部分：分布式文件存储应用接口》（GB/T 31916.3-2018）、《信息技术 大数据存储与处理系统功能要求》（GB/T 37722-2019）、《信息技术 大数据存储与处理系统功能测试要求》（GB/T 38676-2020）等，对

数据存储与管理、应用接口规定、系统功能要求进行了规范约束。

2019年，国家标准《信息技术 大数据存储与处理系统功能要求》（GBT 37722-2019）发布，对大数据存储与处理系统的设计、开发和应用部署发挥了一定指导作用。但是，由于不同行业数据具有领域特色，尤其是农业领域数据多源异构性特点明显，该通用性标准，在指导农业农村领域大数据存储应用方面尚存在一些不足。一是农业农村数据来源多样、数据类型众多，包括调查统计数据、业务系统数据以及遥感、传感器、智能终端等技术装备自动产生数据，各类数据时空维度、体量、更新周期的不同，对储存系统功能及性能要求不尽相同，通用标准因为领域针对性不足，标准缺乏可操作性。二是系统性能是衡量一个系统重要指标，存储系统功能与系统性能密不可分，二者应该统筹考虑。因此，有必要针对农业农村大数据特点，编制《农业农村大数据存储系统功能及性能要求》行业标准，指导不同层级、不同品种农业农村大数据中心建设，提高数据存储系统效率和质量。

（三）起草过程

1. 起草阶段

标准起草单位按照立项计划于2022年4月成立起草组，制订项目实施方案，确定标准主体内容并进行充分论证。

（1）起草组组成：主要起草单位包括XXX。

主要起草人员为：XXX。

具体分工情况如下：

（2）标准主要内容确定和论证过程。2022年4月-2022年6月，对已有相关的农业信息技术、数据存储等方面标准进行充分梳理和理解，结合实际项目需求和业务调研，确定标准规范建设内容。2022年6月-2024年5月，组织专业的标准研编小组人员进行标准规范起草，对于标准规范中有重大分歧意见的内容或部分，充分讨论或咨询相关行业专家意见和建议，形成《农业农村大数据存储系统功能及性能要求》（讨论稿），并注意与国家有关标准的充分衔接。

2. 征求意见阶段

计划分两步实施征求意见：一是对《农业农村大数据存储系统功能及性能要求》（征求意见稿）广泛征求农业农村大数据领域专家、地方农业农村大数据管

理部门、农业农村信息龙头企业的意见和建议；二是对《农业农村大数据存储系统功能及性能要求》（征求意见稿）在大数据中心门户网站公开后征求社会各界征求意见。

将征求到的意见和建议进行整理和消化吸收，与专家充分沟通，形成标准规范《农业农村大数据存储系统功能及性能要求》（送审稿）。

3. 审查阶段

4. 报批阶段（未报批的不写本部分）

二、标准编制原则、主要内容及其确定依据

（一）编制原则

在标准起草过程中，工作组充分考虑了我国农业农村大数据存储系统的建设及管理现状、各省市的农业农村大数据建设水平，重视调查研究，广泛听取各方意见，对我国各地各地农业农村大数据平台的存储系统给出了功能和性能要求，充分保证了标准的科学性、先进性和可操作性、适用性和统一性，同时，遵循 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》等相关标准的规定起草。本次标准的编写主要依据以下原则：

1. 先进性和系统性

充分借鉴国内外研究成果，确保先进性。考虑了各省市县、农业主体、系统开发企业等对于农业农村大数据存储系统应用方向和技术进步，注重科技成果的转化，具有先进性；对全国重点数字化建设较为完善的江苏、河南、上海、浙江等省级大数据建设项目进行调研，借鉴其他部委大数据平台建设规范，总结出适用于农业农村大数据系统软件功能性能等经验成果，既鼓励各方利用先进技术、发展先进设备、保障大数据存储和利用效率的提升，又考虑到农业农村领域内以及与其他领域数据互联互通的现状，满足实际应用场景的需求，实现先进性与系统性并重。

2. 兼容性和扩展性

标准的关键指标突出农业农村领域大数据的特点，各类数据、各级数据在存储系统中的重要性及存储方式等行业应用和服务需求，按照农业农村大数据业务管理模式和实际需求，从总体架构、存储系统及各存储子系统等方面兼顾系统功能和性能，确保其与国内外已有相关大数据存储系统功能和性能标准及文件兼容，

保证农业农村大数据存储系统功能和性能的可扩展性,可适应未来阶段部门机构或业务的调整,预留扩展空间,保证可根据需要进行扩展和细化,以适应农业农村数据的变化和更新。

(二) 标准主要内容

本文件共分为7章。其中第1章是范围,第2章是规范性引用文件,第3章是术语和定义,第4章是缩略语,第5章是系统架构,第6章是功能要求,第7章是性能要求。主要技术内容如下:

(1) 范围:本文件规定了农业农村大数据存储系统的术语和定义、缩略语、系统架构与一般要求、功能要求、性能要求。本文件适用于农业农村大数据存储系统的设计、开发、系统对接、测试和应用部署。

(2) 规范性引用文件:本文件列出了6个规范性引用文件,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。

(3) 术语和定义:本文件对大数据、农业农村大数据、分布式存储3个术语进行了定义。GB/T 35295和NY/T 4261界定的术语和定义适用于本文件,另本文件对农业农村大数据和分布式存储进行了定义,农业农村大数据的定义参考了GB/T 37973中的大数据的定义,同时突出了农业农村数据特点。

(4) 缩略语:对本文件中出现的15个缩略语进行展开和说明。

(5) 系统架构:参考GB/T 37732-2019中云存储系统的架构,在此基础上结合农业农村大数据特征和应用,给出了存储资源层、存储系统层、接口层、应用层四个层级的架构图,并对四个层级的一般功能与性能提出了具体要求。存储系统功能和性能都围绕架构图进行展开。对存储系统的四个层级提出一般功能与性能要求。以表格形式按数据重要性和访问频率分别给出各种数据等级对应的安全性、可用性和响应时间要求,数据存储安全与可用性随数据重要性降低而降低,相应时间随访问频率降低则相应时间可越长。

(6) 功能要求:包括大数据存储功能一般要求和不同数据类型存储要求两大方面。其中,一般要求主要包括:存储资源要求、数据一致性和可靠性要求、兼容性和开放性要求、数据安全性要求、管理和维护要求、能效和成本要求;不同数据类型存储要求主要包括:文件存储功能、对象存储功能、关系型数据存储

功能、空间数据存储功能、图数据存储功能、列式数据存储功能、时序数据存储功能、向量数据存储功能，分别对以上数据存储功能进行了要求。

(7) 性能要求：主要包括存储系统性能指标及测试方法、不同正负等级和不同访问频率数据的存储要求。

(二) 主要内容确定依据

主要技术内容在参考 GB/T 37722-2019 的基础上进行扩展，结合在涉农生产、经营、管理、服务过程中产生的涉及农业农村资源、主体、产品的农业农村大数据特点，在考察浙江等地已建成的农业农村大数据存储系统实际应用成果后，考虑了存储系统发展方向，同时也参考了部分地方标准如山东省地方标准 DB/T37 4529-2022 海洋大数据分级存储技术规范等标准的技术内容，确定了本标准的技术要求。

三、试验验证的分析、综述报告，技术经济论证，预期的经济效益、社会效益和生态效益

(一) 试验或验证的分析、综述报告

1. 试验、验证

为验证《农业农村大数据存储系统功能及性能要求》的完整性和可行性，我们依托大数据中心组织研发的农业农村大数据公共平台基座和全农码平台。在此基础上，我们精选了涵盖农村承包地、“两区”“划定、集体资产、家庭农场等多类业务数据，以确保覆盖全国范围的数据量。这些数据包括空间数据、图数据、结构化数据以及列式存储等多种类型，我们对系统的存储功能和性能进行了广泛的试验。这一系列实验旨在验证本文件所述系统设计的完整性和实施的可行性。

2. 实施案例与落地效果

参考《农业农村大数据存储系统功能及性能要求》的核心内容，目前我们已经建立了农业农村大数据公共平台基座和全农码平台。这些平台充分利用了大数据存储平台，涵盖多种数据存储方式，包括关系数据、空间数据、图数据、列式数据以及对象与文件存储。

在关系型数据存储方面，大数据存储平台目前存储了 143GB 的数据，包含超过 25 亿条与全农码业务相关的数据。在空间数据存储中，平台目前存储了超过 2GB 的空间数据，这些数据分布在大约 100 张表中，总计约 350 万条数据，

为复杂的空间分析和地理信息处理提供了坚实的数据基础。在图数据存储中，平台成功存储了 27 亿个点和 97 亿条边，达到了高效的存储和管理水平。在列式数据存储中，平台目前存储了 6.5TB 的数据，这些数据分布在 607 张表中，总计约 357 亿条数据，为大规模数据分析提供了有力支持。在对象与文件存储中，平台目前存储了约 20GB 的数据，确保了多种类型文件和对象的安全存储和高效检索。

在全农码平台中，大数据存储平台的使用覆盖了多个方面，平台提供了强大的数据查询功能，利用大数据存储平台的索引和查询优化技术，实现了海量数据的快速检索。无论是简单的关系型数据查询，还是复杂的图数据、空间数据查询，系统都能在秒级响应时间内返回结果。大数据存储平台为全农码平台提供了丰富的数据分析能力，支持实时分析和批量处理。平台集成了多种分析工具和算法，能够进行复杂的数据挖掘、统计分析和预测模型的训练与应用，为决策支持提供有力依据。在对象与文件存储中，平台确保了各类文件和对象的安全存储与高效管理。系统具备强大的权限控制和数据备份功能，保证了数据的安全性和完整性。通过列式存储和并行处理技术，平台在处理大规模数据时表现出色。大数据存储平台的优化策略和高效算法使得平台在高并发访问和大量数据处理情况下依然能够保持高性能。

经过两年多的实际业务验证，全农码平台运行稳定，性能优越，完全满足业务需求。这些成果不仅验证了《农业农村大数据存储系统功能及性能要求》的完整性和可行性，也为未来的大数据应用和发展提供了重要参考和保障。

（二）技术经济论证

《农业农村大数据存储系统功能及性能要求》的编制结合农业农村大数据和业务情况，充分借鉴国际、国外、国内等相关标准的研究成果，广泛征求农业农村大数据领域专家、管理部门和相关企业的意见和建议，详尽考虑农业农村大数据存储实施所需条件、投入、成本、难易程度，具备技术先进性和经济合理性。

1.技术先进性

ISO/IEC 27033-1 是国际标准化组织（ISO）和国际电工委员会（IEC）发布的关于信息安全指南的标准之一，涉及到大数据的安全管理，与之相比，《农业农村大数据存储系统功能及性能要求》更侧重于农业领域的特定需求，如数据的多样性和复杂性。ISO/IEC 20943 是关于大数据概念、参考架构和功能要求的国

际标准，《农业农村大数据存储系统功能及性能要求》在参考这些国际标准的基础上，针对农业农村领域的特殊性进行定制和优化。美国国家标准与技术研究院（NIST）发布的大数据框架提供一套全面的大数据管理和分析指南，《农业农村大数据存储系统功能及性能要求》在借鉴这些框架的基础上，结合中国农业农村的实际情况进行了本土化调整。欧盟数据保护条例（GDPR）主要关注个人数据保护，对数据安全性和隐私保护的要求对大数据存储系统的设计有重要影响，GB/T 37722-2019 是中国关于大数据存储与处理系统功能要求的国家标准，《农业农村大数据存储系统功能及性能要求》在遵循这一标准的基础上，针对农业农村领域的特点进行更详细的规定。GB/T 35295 是关于大数据术语的国家标准，为农业农村大数据标准提供了术语定义的基础。《农业农村大数据存储系统功能及性能要求》同样强调了数据安全和隐私保护，确保符合国际标准，且对数据存储的资源要求、系统功能要求、性能要求进行了创新性的设计，在技术上具有引领性和先进性。

2.实施合理性

《农业农村大数据存储系统功能及性能要求》针对农业领域的特定需求进行了定制化设计，如对多源异构数据的存储和管理，这在国际上是领先的实践。与国际上在数据安全和隐私保护方面的标准保持一致，甚至在某些方面更加严格，以适应中国农业农村的特殊环境。整合了最新的大数据技术，如分布式存储、云计算、人工智能等，确保技术的先进性。提出了具体的性能指标和测试方法，确保大数据存储系统能够满足农业农村领域的实际需求。考虑未来技术发展的趋势，确保系统的可扩展性，能够适应未来农业农村大数据的发展。

通过以上比较分析，《农业农村大数据存储系统功能及性能要求》标准在技术内容上处于较高水平，既符合国内外标准的要求，又具有针对中国农业农村领域的先进性和适用性。

3.经济合理性

《农业农村大数据存储系统功能及性能要求》的编写详尽评估农业农村大数据存储实施条件、投入、成本、难易程度。评估当前农业农村大数据存储系统的技术水平，包括硬件设备、软件系统、网络基础设施等，确定升级或新建系统所需的技术条件，实施标准所需的人员培训，包括数据管理、系统维护、数据分析

等方面的专业技能培训，政府对农业农村大数据发展的政策支持力度，包括资金支持、税收优惠、政策引导等。评估实施标准所需的硬件设备投入，如服务器、存储设备、网络设备等，所需的软件系统投入，包括数据库管理系统、数据分析工具、数据安全软件等。所需的人员投入，包括数据管理人员、系统维护人员、数据分析人员等。评估该标准的直接成本，包括硬件设备、软件系统、人员培训等直接投入成本，间接成本，包括系统维护、数据安全、技术支持等长期运营成本，机会成本，包括不实施该标准可能带来的潜在损失，如数据丢失、效率低下等。评估实施标准的技术难度，包括技术升级、系统集成、数据迁移等，操作难度，包括人员培训、流程调整、系统维护等，可能面临的风险，如技术更新换代、市场需求变化等。

充分考虑实施该标准所需的成本（如硬件设备、软件系统、人员培训等）与预期的经济效益（如提高数据存储效率、降低数据丢失风险、提升数据分析能力等）、社会效益（如提高农业农村信息化水平、促进乡村振兴战略的实施等）、生态效益（如通过优化数据存储减少能源消耗、提高资源利用率等）之间的关系，合理评估标准可能面临的风险（如技术更新换代、市场需求变化等），制定实施该标准的策略和步骤，如分阶段实施、优先支持关键领域等，确保经济上的可行性。

通过上述综合分析，可以全面论证《农业农村大数据存储系统功能及性能要求》标准的经济合理性，为标准的推广和实施提供有力的支持。

（三）预期的经济效益、社会效益和生态效益

本标准对推进各地农业农村大数据项目的稳定实施具有重要价值，推进农业农村数据资源利用方式转变，为各级农业农村大数据发展提供可靠存储系统功能及性能，极大地方便了农业农村大数据的共享，减少重复性摸索建设，具有极大的经济效益。同时本标准提高了数据存储系统效率和质量，简化数据的存储和共享使用，促进了农业农村大数据与其他数据间的交互，带来了社会效益。标准在系统功能和性能与大数据使用的灵活匹配之间，更加注重绿色化能效和控制存储成本，具有极大的生态效益。

四、与国际、国外同类标准技术内容的对比情况，或者与测试的国外样品、样机的有关数据对比情况

国际大数据相关标准的研制主要集中在国际标准化组织/国际电工技术委员会第一技术委员会第9大数据工作组（ISO/IEC JTC1/WG9，以下简称WG9）以及国际电信联盟电信标准分局（以下简称ITU-T）。

WG9发布的《信息技术--大数据参考体系结构》系列标准，为大数据提供了一个全面的参考框架，包括框架和应用过程、用例和派生需求、参考体系架构、安全和隐私、标准路线图等部分。这些标准为大数据技术的应用和实施提供了基础性指导。

ITU-T发布的大数据相关系列标准，如Y.3601《大数据-框架和需求数据交换》等，侧重于大数据交换的框架和需求，功能架构，数据来源要求等方面。这些标准在数据交换和互操作性方面提供了具体的技术指导。

美国也较系统性地开展了大数据标准化的工作。美国国家标准与技术研究院属下的大数据公共工作组（NBD-PWG）发布的《大数据互操作框架》系列标准提供了大数据互操作性的定义、分类法、用例、安全和隐私等方面的指导。其中，第6卷《大数据参考架构》为大数据技术的设计和 implement 提供了综合性架构。这些标准在大数据互操作性方面提供了较为全面的框架。

国际标准，并未专门针对农业农村大数据存储系统进行详细规定。本标准，在参考上述国际和国外标准的基础上，针对农业农村领域的特定需求进行了定制化设计。与现有国际标准相比，本标准具备以下特点：1、满足特定领域的需求（考虑到农业农村大数据的特殊性，如数据的多源异构性、时空维度、体量和更新周期等）；2、更丰富的系统性能的衡量（包括存储系统的可用性、维护性、I/O性能、延迟和吞吐量等关键性能指标）；3、不同数据类型的存储要求（针对农业农村大数据中的文件存储、对象存储、关系型数据存储、空间数据存储、图数据存储、列式数据存储、时序数据存储和向量数据存储等不同类型的数据存储要求）；4、具体的安全性和隐私保护要求（包括数据的加密、访问控制、审计日志、多租户管理和数据隔离等）；5、考虑成本效益（在满足性能和功能需求的同时，考虑存储成本的控制）。

五、以国际标准为基础的起草情况，以及是否合规引用或者采用国际国外标准，并说明未采用国际标准的原因

本标准未引用、采用或参考国际国外标准。本标准是根据我国农业农村大数据发展的现实情况，在总结地方应用经验的基础上制定的，适合我国的国情，具有一定的先进性。

六、与有关法律、行政法规及相关标准的关系

本标准遵循法律、法规，符合国家有关现行法律和强制性国家标准的规定，与相关法律和国家标准相协调，无冲突。

相关现行的法律法规和国家标准如下：

《中华人民共和国数据安全法》；

《中华人民共和国网络安全法》；

《中华人民共和国个人信息保护法》；

GB/T 35295 《信息技术 大数据 术语》；

GB/T 37732 《信息技术 云计算 云存储系统服务接口功能》；

GB/T 37737 《信息技术 云计算 分布式块存储系统总体技术要求》；

GB/T 37973 《信息安全技术 大数据安全管理指南》；

GB/T 38676 《信息技术 大数据 存储与处理系统功能测试要求》；

NY/T 4261 《农业大数据安全管理指南》。

七、重大分歧意见的处理经过和依据

本标准在制定的过程中未出现重大分歧意见。

八、涉及专利的有关说明

无。

九、实施标准的要求，以及组织措施、技术措施、过渡期和实施日期的建议等措施建议

本标准推荐为推荐性行业标准，建议过渡期为一个月。建议过渡期间进行本标准的宣贯培训工作，根据本标准的适用范围，将主要面向各级农业农村机关、直属单位等用户进行标准的培训与宣贯，采用专家讲座、系列课程、交流答疑、发放宣贯材料等方式，积极推进标准的实施应用。

十、其他应当说明的事项
无。