



中华人民共和国国家标准

GB/T 44216—2024

信息技术 大数据 批流融合计算技术要求

Information technology—Big data—Technical requirements for integrated batch
and streaming computing

2024-07-24 发布

2025-02-01 实施

国家市场监督管理总局
国家标准化管理委员会 发布

目 次

| | |
|-------------------------------|-----|
| 前言 | III |
| 引言 | IV |
| 1 范围 | 1 |
| 2 规范性引用文件 | 1 |
| 3 术语和定义 | 1 |
| 4 缩略语 | 2 |
| 5 系统架构 | 2 |
| 6 技术要求 | 3 |
| 6.1 统一资源管理 | 3 |
| 6.2 统一计算框架 | 4 |
| 6.3 统一 SQL 接口 | 4 |
| 6.4 统一 API | 5 |
| 6.5 统一作业管理 | 5 |
| 6.6 统一权限管理 | 6 |
| 7 扩展性要求 | 6 |
| 8 兼容性要求 | 7 |
| 9 性能指标 | 7 |
| 附录 A（资料性） 批流融合计算应用场景 | 8 |
| A.1 金融行业 | 8 |
| A.2 智能制造行业 | 8 |
| A.3 物联网 | 8 |
| A.4 航空航天 | 8 |
| 附录 B（资料性） 字符类型及操作中英文对照表 | 9 |
| 参考文献 | 11 |

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。

本文件由全国信息技术标准化技术委员会（SAC/TC 28）提出并归口。

本文件起草单位：阿里云计算有限公司、中国电子技术标准化研究院、浪潮电子信息产业股份有限公司、华为技术有限公司、浙江中烟工业有限责任公司、蚂蚁科技集团股份有限公司、腾讯云计算（北京）有限责任公司、杭州海康威视数字技术股份有限公司、北京易华录信息技术股份有限公司、上海宝信软件股份有限公司、北京柏睿数据技术股份有限公司、山东浪潮数据库技术有限公司、上海计算机软件技术开发中心、中国铁道科学研究院集团有限公司、中国电子系统技术有限公司、浪潮软件科技有限公司、中通服咨询设计研究院有限公司、网易（杭州）网络有限公司、中国南方电网有限责任公司、北京百分点科技集团股份有限公司、广东电网有限责任公司、北京华胜天成科技股份有限公司、浙江邦盛科技股份有限公司、中国石油化工集团有限公司、普元信息技术股份有限公司、方正国际软件（北京）有限公司、浙江创邻科技有限公司、四川发展数字金沙科技有限公司、北京庚顿数据科技有限公司、陕西省信息化工程研究院、成都数据集团股份有限公司、国能信息技术有限公司、成都数之联科技股份有限公司、云上贵州大数据产业发展有限公司、药融云数字科技（成都）有限公司、浙江智奥科技有限公司、深圳市中丝贸科技有限公司、深圳亿维锐创科技股份有限公司、河南省新星科技有限公司、浙江方信标准技术有限公司、海南电网有限责任公司、中国电信集团财务有限公司、复旦大学、上海交通大学。

本文件主要起草人：朱松、陈守元、许洁、王峰、吴涛、刘健、杨锐、金泳、郭智慧、刘海涛、蒋楠、黄超、陈小龙、张延生、王凌、黄明、赵菁华、朴晟宏、高阳、张黎明、陈敏刚、刘陈宇、吴艳华、刘国栋、杨旭、王金超、张永良、马进、王宇静、陈彬、杨秋勇、钱正浩、梁盈威、孙伟、梁钢、王新根、臧一超、张煜、周研、刘宇峰、王晋晖、张勇、李傲铁、李正、段智琛、傅彦、周俊临、黄明峰、刘军、王军刚、张晖、李明、王中健、周小华、丘琳、许宏安、皮志新、董卫魏、郭威、张磊、冯曹冲、陈勇锦、王毅、李爽、郑忠斌。

引 言

随着数据量的增长，分布式计算模式逐渐成为大数据处理和计算的主流架构。为了满足程序的健壮性和实时性要求，设计出了许多分布式计算框架，以屏蔽底层复杂的任务划分和集群调度细节，其中最常用的两种是批计算技术和流式计算技术，二者有着迥异的编程模型和编程接口，适用于不同的计算场景。批计算技术适合需访问全套记录才能完成的计算工作，流计算技术很适合用来处理需对变动或峰值做出响应，并且关注一段时间内变化趋势的数据。

在实际应用中，经常会遇到两种计算技术共同工作的情况。将两种计算框架进行简单的叠加，则需要在两个不同的引擎上实现相同的执行逻辑，还需要手工合并不同引擎的输出结果。如果需要更改查询逻辑，两个系统也需要同时进行改动。这会极大地增加工程的开发和维护成本。因此，统一的批流融合计算技术成为了大数据领域的重要发展趋势。

2 信息技术 大数据 批流融合计算技术要求

1 范围

本文件规定了大数据批流融合计算技术要求，包括：技术要求、扩展性要求、兼容性要求、性能指标等。

本文件适用于批流融合计算系统的设计、开发和部署。用户理解、采用、建设批流融合计算技术和批流融合产品与服务评估参照执行。

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。

GB/T 35295—2017 信息技术 大数据 术语

3 术语和定义

GB/T 35295—2017 界定的以及下列术语和定义适用于本文件。

3.1

批处理 batch processing

将一个大型作业分解成为多个任务交由多个节点分别处理，再将分解后多个任务处理的结果汇总起来得出最终的分析结果的计算框架。

3.2

流处理 stream processing

针对处理高速并发且时效性有较高要求的大规模计算场景，能够对具有实时、高速、瞬时性等特性的流式数据进行实时处理的计算框架。

3.3

批流融合计算 integrated batch and streaming computing

能够同时支持批处理和流处理等的计算框架。

3.4

分散-聚集 map-reduce

所需的计算被划分并分布在多个节点上，整体结果由每个节点的结果合并而成的大数据集的处理形式。

示例：MapReduce（包含 Map 和 Reduce 两个计算过程的一种计算模型）就是采用分散-聚集的处理形式。

3.5

租户 tenant

对一组物理和虚拟资源进行共享访问的一个或多个云服务用户。

3.6

多租户 multi-tenancy

通过对物理或虚拟资源的分配保证多个租户以及他们的计算和数据彼此隔离和不可访问。