



中华人民共和国国家标准

GB/T 41818—2022

信息技术 大数据 面向分析的 数据存储与检索技术要求

Information technology—Big data—Technical requirements for
analysis-oriented data storage and retrieval

2022-10-12 发布

2023-05-01 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
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	1
5 概述	2
6 数据表结构	2
6.1 数据表	2
6.2 子表	2
6.3 行组	3
6.4 列组	3
6.5 数据页	4
7 索引技术要求	4
8 数据存储技术要求	4
8.1 基本要求	4
8.2 存储布局实现	4
8.3 数据处理	5
9 数据检索技术要求	6
附录 A (资料性) 面向分析的数据存储与检索应用示例	7
参考文献	9

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：华为技术有限公司、北京大学、中国电子技术标准化研究院、中国人民大学、北京航天智造科技发展有限公司、中汽研汽车检验中心(天津)有限公司、成都中科大旗软件股份有限公司、浪潮软件科技有限公司、江苏赛西科技发展有限公司、深圳赛西信息技术有限公司、浪潮电子信息产业股份有限公司、北京易华录信息技术股份有限公司、杭州中奥科技有限公司、方正国际软件(北京)有限公司、山东黄河河务局山东黄河信息中心、山东省计算中心(国家超级计算济南中心)、四川大学华西医院、湖南财政经济学院、杭州市第七人民医院、北京工业大学、北京理工大学、中山大学、桂林电子科技大学、天津大学、中冶赛迪工程技术股份有限公司、北京能科瑞元数字技术有限公司、北京启迪数字科技集团有限公司、中国人民解放军国防科技大学、西北工业大学、西安电子科技大学、腾讯科技(深圳)有限公司、河南云政数据管理有限公司、特斯联科技集团有限公司、深圳龙岗智能视听研究院、星环信息科技(上海)有限公司。

本文件主要起草人：梅宏、杜小勇、符海芳、陈亮、范科峰、张群、赵华、赵俊峰、王亚沙、刘驰、尹卓、王为中、曹幼林、卞昊穹、金国栋、陈跃国、李民东、刘哲、邹萍、殷晋、贺可勋、冯谦、李武鸿、郑申俊、李冰、张亮、石征、袁玥、赵斌、李威、王树良、洪江、黄先芝、胡清、王凌、陶智敏、李振东、张煜、逢锦山、张媛、潘彬、孙光、章俊航、杨绍武、史殿习、杨震、于海阳、万海、何倩、雷建军、李斐、潘兆庆、彭勃、刘国杰、张超超、贾晓杰、沈丽丽、张星星、邓乔、黎方学、车伟伟、肖学文、毛尚伟、杨刚、姚远、杨钰、张大鹏、杨洪山。

信息技术 大数据 面向分析的 数据存储与检索技术要求

1 范围

本文件规定了支撑多应用融合分析的大数据列式存储与检索技术要求。
本文件适用于面向分析的数据存储与检索的系统或子系统的设计、开发和使用。

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。

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

元数据 metadata

定义和描述其他数据的数据。

[来源:GB/T 18391.1—2009,3.2.16]

3.2

存储布局 storage layout

数据在存储系统中的存储规划和安排。

注:通常包括各数据单元在存储系统中相对的存储顺序和存储位置的偏移等。

3.3

检索 retrieval

使用一定方法或工具,从数据集合中找出用户所需数据或数据相关信息的过程。

4 缩略语

下列缩略语适用于本文件。

CSV:逗号分隔值(Comma Separated Values)

EB:百亿亿字节(Exa Byte)

JSON:脚本对象(JavaScript Object Notation)

OLAP:联机分析处理(On-line Analytical Processing)

PB:千万亿字节(Peta Byte)

SQL:结构化查询语言(Structured Query Language)